

MENU

SEARCH

INDEX

DETAIL

1/1



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11)Publication number: 10111761

(43)Date of publication of application: 28.04.1998

(51)Int.Cl.

G06F 3/06
G11B 20/10
H04L 29/08
H04N 5/78
H04N 7/16

(21)Application number: 08278359

(71)Applicant:

MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing: 21.10.1996

(72)Inventor:

ITO YUKIKO
TANAKA TSUTOMU
DOI SHINZO
KOKADO TAKESHI
KARIBE TOMOYUKI
TAMAI MASARO

(30)Priority

Priority number: 07278799 Priority date: 26.10.1995 Priority country: JP

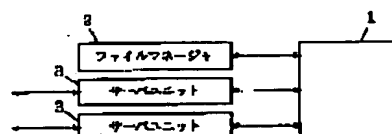
07278813	26.10.1995	JP
07278814	26.10.1995	JP
08125146	20.05.1996	JP
08213556	13.08.1996	JP

(54) FILE SYSTEM

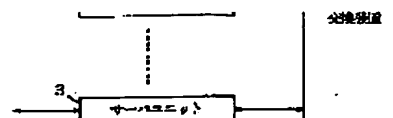
(57)Abstract:

PROBLEM TO BE SOLVED: To provide the data in a short response time by receiving a get resource message and then updating and recording the processing use frequency when a requested band is secured.

SOLUTION: The message interface part of a file manager 2 identifies the messages which are transferred among server



units 3 via a switching device 1 and then instructs the processing to the function contained in the manager 2. A resource management part manages the capacity and the use frequency of the read band of a storage and the input/output band of the device 1 in a file system. That is, the resource management part receives a get resource message that previously requests the security of a necessary band when the data are read or written and decides whether the band requested by the get resource message can be secured based on the maximum value and the use frequency of the processing capability when the get resource message is received. If the requested band is secured, the use frequency of the band is updated and recorded.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998 Japanese Patent Office

[MENU](#)

[SEARCH](#)

[INDEX](#)

[DETAIL](#)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-111761

(43) 公開日 平成10年(1998) 4月28日

(51) Int.Cl. ⁸	識別記号	F I
G 0 6 F 3/06	3 0 1	G 0 6 F 3/06 3 0 1 M
G 1 1 B 20/10		G 1 1 B 20/10 D
H 0 4 L 29/08		H 0 4 N 5/78 B
H 0 4 N 5/78		7/16 A
7/16		H 0 4 L 13/00 3 0 7 Z
審査請求 未請求 請求項の数97 O L (全 67 頁)		

(21) 出願番号 特願平8-278359

(22) 出願日 平成8年(1996)10月21日

(31) 優先権主張番号 特願平7-278799

(32) 優先日 平7(1995)10月26日

(33) 優先権主張国 日本 (J P)

(31) 優先権主張番号 特願平7-278813

(32) 優先日 平7(1995)10月26日

(33) 優先権主張国 日本 (J P)

(31) 優先権主張番号 特願平7-278814

(32) 優先日 平7(1995)10月26日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 伊藤 由起子

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 田中 勉

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 土居 晋三

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(74) 代理人 弁理士 小笠原 史朗

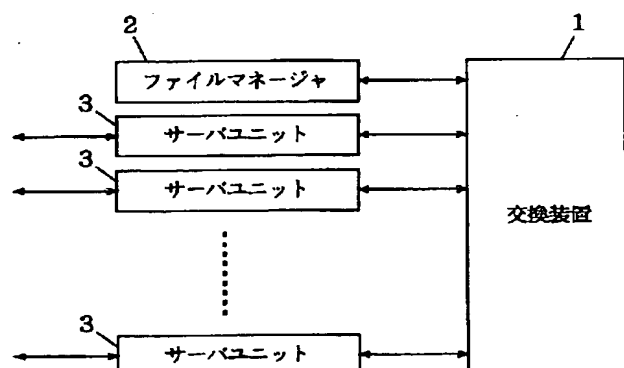
最終頁に続く

(54) 【発明の名称】 ファイルシステム

(57) 【要約】

【目的】 ビデオサーバの使用効率を向上し、ユーザからの要求に対して素早い応答時間で映像を提供し、様々な再生速度を要求するユーザをサポートするファイルシステムを提供する。

【構成】 外部からゲットリソースメッセージが到着すると、リソース割り当て部102は、リソースの予約を行う。データの読み出し、書き込みに際しては、申告値管理部111にてパラメータをチェックし、スロット割り当て部103でリソースの割り当てを行うことにより、データの読み出しに関して、遅延品質を保証した、効率のよいファイルシステムを構築する。



【特許請求の範囲】

【請求項 1】 外部からの要求に応じて、データを記憶する記憶装置に対してデータの読み出しや書き込みを行うファイルシステムであって、読み出しや書き込みの処理能力の最大値と、当該処理能力のうちの使用量とを管理するリソース管理部を備え、前記リソース管理部は、データの読み出しや書き込みを行う際には、予め必要な帯域の確保を要求するゲットリソースメッセージを受信し、前記ゲットリソースメッセージを受信したとき、前記処理能力の最大値および前記使用量に基づいて、当該ゲットリソースメッセージで要求される帯域を確保できるかを判断し、前記判断の結果、要求される帯域を確保できる場合には、前記使用量を更新して記憶することを特徴とする、ファイルシステム。

【請求項 2】 前記ゲットリソースメッセージには、パラメータとして、必要な帯域量を示す情報が付与されていることを特徴とする、請求項 1 に記載のファイルシステム。

【請求項 3】 前記ゲットリソースメッセージのパラメータである、必要な帯域量を示す情報は、前記記憶装置に対して、固定長のデータブロックを読み出しまたは書き込みする間隔とすることを特徴とする、請求項 2 に記載のファイルシステム。

【請求項 4】 前記ゲットリソースメッセージの要求に従って確保する帯域量は、前記リソース管理部において予め記憶されていることを特徴とする、請求項 1 に記載のファイルシステム。

【請求項 5】 前記ゲットリソースメッセージには、読み出し用、書き込み用を識別する情報が付与されており、前記リソース管理部は、前記ゲットリソースメッセージの要求に従って確保する帯域量として、読み出し用の帯域量と書き込み用の帯域量との 2 種類を記憶していることを特徴とする、請求項 4 に記載のファイルシステム。

【請求項 6】 前記ゲットリソースメッセージには、パラメータとして、読み出しの要求を行ってからデータの送信が開始されるまでの第 1 の遅延時間が付与されており、前記リソース管理部は、予めファイルシステムが保証可能な遅延時間の最短値を記憶しており、前記ゲットリソースメッセージを受信したとき、前記第 1 の遅延時間と前記最短値とを比較し、前記比較の結果、前記第 1 の遅延時間が前記最短値よりも短い場合、帯域の確保を行わないことを特徴とする、請求項 1 に記載のファイルシステム。

【請求項 7】 外部から到着するメッセージに付与され

ているパラメータをチェックする申告値管理部をさらに備え、

前記申告値管理部は、

データの読み出しを行う際には、読み出されたデータの送信が開始されるまでの第 2 の遅延時間がパラメータとして付与されたリードブロックメッセージを受信し、前記リードブロックメッセージを受信したとき、前記第 2 の遅延時間と前記第 1 の遅延時間とを比較し、前記比較の結果、前記第 2 の遅延時間のほうが短いと判断した場合、前記リードブロックメッセージの送信元に対して違反通知を送信することを決定することを特徴とする、請求項 6 に記載のファイルシステム。

【請求項 8】 外部から到着するメッセージに付与されているパラメータをチェックする申告値管理部をさらに備え、

前記申告値管理部は、

ファイルシステムが保証可能な遅延時間の最短値を記憶しており、

データの読み出しを行う際には、読み出されたデータの送信が開始されるまでの第 2 の遅延時間がパラメータとして付与されたリードブロックメッセージを受信し、前記リードブロックメッセージを受信したとき、前記第 2 の遅延時間と前記最短値とを比較し、前記比較の結果、前記第 2 の遅延時間のほうが短いと判断した場合、前記リードブロックメッセージの送信元に対して違反通知を送信することを決定することを特徴とする、請求項 1 に記載のファイルシステム。

【請求項 9】 前記ゲットリソースメッセージには、パラメータとして、読み出しの要求を行ってからデータの送信が開始されるまでの遅延時間が付与されており、前記リソース管理部は、

前記ゲットリソースメッセージを受信したとき、前記遅延時間を満たすために必要な帯域量を算出し、前記算出の結果得た帯域量と前記ゲットリソースメッセージの要求に従って確保すべき帯域量とを比較して、大きい方の帯域量を確保できるかどうかを判断することを特徴とする、請求項 1 に記載のファイルシステム。

【請求項 10】 外部から到着するゲットリソースメッセージに付与されているパラメータをチェックする申告値管理部をさらに備え、

前記申告値管理部は、データの書き込み時において、

データ到着の速度を観測し、

前記観測の結果得たデータ到着の速度と、前記ゲットリソースメッセージ受付時に確保した帯域量とを比較し、前記比較の結果、前記データ到着の速度のほうが大きいと判断したときは、データ書き込みは行わず、前記ゲットリソースメッセージ送信元に対して違反通知を行うことを特徴とする、請求項 1 に記載のファイルシステム。

【請求項 11】 外部から到着するメッセージに付与されているパラメータをチェックする申告値管理部をさら

に備え、
 前記申告値管理部は、
 データの読み出し時には、単位時間当たりの読み出しデータ量がパラメータとして付与されたリードブロックメッセージを受信し、
 前記リードブロックメッセージを受信したとき、前記単位時間当たりのデータ量からデータ読み出し速度を算出し、
 前記算出によって得たデータ読み出し速度と、前記ゲットリソースメッセージ受付時に確保した帯域量とを比較し、
 前記比較の結果、前記データ読み出し速度のほうが大きい場合、データの読み出しを中止し、前記リードブロックメッセージ送信元に対して違反通知を送信することを決定することを特徴とする、請求項1に記載のファイルシステム。

【請求項12】 前記処理能力の最大値および使用量は、単位時間当たりの読み出しデータ量であることを特徴とする、請求項1に記載のファイルシステム。

【請求項13】 前記ゲットリソースメッセージには、パラメータとして、読み出し、書き込みを区別するフラグと、必要な帯域量を示す情報とが付与されており、
 前記リソース管理部は、前記ゲットリソースメッセージを受信したとき、前記フラグを参照し、読み出しの場合には、前記ゲットリソースメッセージに付与されているパラメータによって決定される帯域量を確保すべき帯域量とし、書き込みの場合には、前記ゲットリソースメッセージに付与されているパラメータによって決定される帯域量に一定値を乗じたものを確保すべき帯域量とすることを特徴とする、請求項12に記載のファイルシステム。

【請求項14】 前記処理能力の最大値および使用量は、単位時間当たりの書き込みデータ量であることを特徴とする、請求項1に記載のファイルシステム。

【請求項15】 前記ゲットリソースメッセージには、パラメータとして、読み出し、書き込みを区別するフラグと、必要な帯域量を示す情報とが付与されており、
 前記リソース管理部は、前記ゲットリソースメッセージを受信したとき、前記フラグを参照し、書き込みの場合には、前記ゲットリソースメッセージに付与されているパラメータによって決定される帯域量を確保すべき帯域量とし、読み出しの場合には、前記ゲットリソースメッセージに付与されているパラメータによって決定される帯域量に一定値を乗じたものを確保すべき帯域量とすることを特徴とする、請求項14に記載のファイルシステム。

【請求項16】 データを格納した複数のブロック格納ユニットと、ユーザからの要求に応じてデータの書き込みや読み出しを制御する単数または複数のI/Oユニットと、各ブロック格納ユニットおよび各I/Oユニット

の間を接続する交換装置とを備え、各I/Oユニットは、ユーザから要求されたデータの大きさを固定長のブロックに分割し、ブロック毎にブロック格納ユニットに対してブロック読み出し要求を送出し、ブロックの読み出しを行うファイルシステムにおいて、

前記I/Oユニットは、
 一定時間毎に時間を区切ったMS到着スロットを設定し、当該MS到着スロットには、一定数のブロック読み出し要求を登録できるものとし、
 前記ブロック読み出し要求送出の際には、当該ブロック読み出し要求に対して前記MS到着スロットを割り当て、

前記割り当てたMS到着スロットに対応した時刻に基づき、前記ブロック格納ユニットから前記交換装置への入力が許可される時刻を算出し、

前記算出された時刻と現在時刻との差の時間を遅延限界値として前記ブロック読み出し要求に付与して前記ブロック格納ユニット宛に送出し、それによって前記ブロック読み出し要求に回答して前記ブロック格納ユニットから読み出されるブロックは、当該ブロック読み出し要求が当該ブロック格納ユニットに到着してから前記遅延限界値の時間の後に、前記交換装置に対する入力が許可されることを特徴とする、ファイルシステム。

【請求項17】 その内部に前記I/Oユニットおよび前記ブロック格納ユニットが設けられたサーバユニットを複数組備えた、請求項16に記載のファイルシステム。

【請求項18】 前記I/Oユニットは、前記ブロック読み出し要求の内、前記交換装置に入力されないブロック読み出し要求に対しては、前記MS到着スロットの割り当てを行わないことを特徴とする、請求項16に記載のファイルシステム。

【請求項19】 読み出しや書き込みの処理能力の最大値と、当該処理能力のうちの使用量とを管理するリソース管理部をさらに備え、

前記リソース管理部は、
 データの読み出しや書き込みを行う際には、予め必要な帯域の確保を要求するゲットリソースメッセージを受信し、

前記ゲットリソースメッセージを受信したとき、前記処理能力の最大値および前記使用量に基づいて、当該ゲットリソースメッセージで要求される帯域を確保できるかを判断し、

前記判断の結果、要求される帯域を確保できる場合には、前記使用量を更新して記憶し、

前記I/Oユニットは、
 ファイルの読み出しに必要な帯域を確保している場合には、確保した帯域だけのデータを読み出すために必要となる前記ブロック読み出し要求の発行間隔を T_{min} に設定し、当該 T_{min} の間隔で、前記MS到着スロ

ットの期待スロットを決定し、
前記決定した期待スロットに空きが存在すれば、当該期待スロットを予約し、

前記決定した期待スロットに空きが存在しない場合には、順次、次の時刻のMS到着スロットを参照して、空きが存在するMS到着スロットを予約し、
データ読み出しの際には、前記予約された前記MS到着スロットに前記ブロック読み出し要求の割り当てを行うことを特徴とする、請求項16に記載のファイルシステム。

【請求項20】 前記I/Oユニットは、前記MS到着スロットの予約を、確保された帯域が大きいものから小さいものの順に行うことを特徴とする、請求項19に記載のファイルシステム。

【請求項21】 前記I/Oユニットは、前記MS到着スロットの予約を、次の期待スロットが到来するタイミングが早いものから順番に行うことを特徴とする、請求項19に記載のファイルシステム。

【請求項22】 前記I/Oユニットは、前記MS到着スロットの割り当てを行った結果、割り当てたMS到着スロットよりも前の時刻で、かつ同一ユーザのために予約されたMS到着スロットが存在する場合には、当該予約されたMS到着スロットを割り当てがされなかったMS到着スロットとし、リソースの獲得を行っていないユーザからのデータの読み出し要求に対し、当該割り当てがされなかったMS到着スロットを割り当ててことを特徴とする、請求項19に記載のファイルシステム。

【請求項23】 前記I/Oユニットは、前記MS到着スロットの予約をゲットリソースメッセージを受け付けているユーザすべてについて行った結果、予約が行われなかった前記MS到着スロットに対しては、リソースの獲得を行っていないユーザからのデータの読み出し要求に割り当ててことを特徴とする、請求項19に記載のファイルシステム。

【請求項24】 データを格納した単数または複数のブロック格納ユニットと、
ユーザからの要求に応じてデータの書き込みや読み出しを制御する単数または複数のI/Oユニットと、
各前記ブロック格納ユニットおよび各前記I/Oユニットの間を接続する交換装置とをさらに備え、
前記I/Oユニットは、
ユーザから書き込みを要求されたデータの大きさを固定長のブロックに分割し、当該ブロック毎に前記ブロック格納ユニットに対してブロック読み出し要求を送出することにより、ブロックの読み出しを行い、
前記ブロック格納ユニットに対応させて、一定時間毎に区切られた書き込み用スロットを設定し、ブロック書き込み要求を送出する際には、当該書き込み用スロットを割り当てて送することを特徴とする、請求項1に記載のファイルシステム。

【請求項25】 複数の記憶装置に対してファイルデータの入出力を行うファイルシステムにおいて、ファイルデータを各記憶装置に分散配置する方法であって、
各前記記憶装置が有する帯域に比例してファイルデータを配置する比率を決定し、当該決定された比率に従ってファイルデータを各記憶装置に分散して格納することを特徴とする、ファイルデータの分散配置方法。

【請求項26】 複数の記憶装置に対してファイルデータの入出力を行うファイルシステムにおいて、ファイルデータを各記憶装置に分散配置する方法であって、
各前記記憶装置が有する記憶容量に比例してファイルデータを配置する比率を決定し、当該決定された比率に従ってファイルデータを各記憶装置に分散して格納することを特徴とする、ファイルデータの分散配置方法。

【請求項27】 複数の記憶装置に対してファイルデータの入出力を行うファイルシステムにおいて、ファイルデータを各記憶装置に分散配置する方法であって、
各前記記憶装置が有する帯域および記憶容量に比例して、ファイルデータを配置する比率を決定し、当該決定された比率に従ってファイルデータを各記憶装置に分散して格納することを特徴とする、ファイルデータの分散配置方法。

【請求項28】 複数の記憶装置に対してファイルデータの入出力を行うファイルシステムにおいて、各記憶装置に対するファイルデータの入出力帯域を管理する装置であって、
各前記記憶装置が有する帯域を管理する帯域管理部と、
各前記記憶装置が有する記憶容量を管理する容量管理部と、
予め規定されたアルゴリズムに従って、各記憶装置にファイルデータを分配する比率を決定する分配比率決定部と、
前記帯域管理部が保持する管理情報と、前記分配比率決定部で決定された分配の比率とに基づいて、全記憶装置をまとめた場合の帯域を決定する帯域計算部と、
前記容量管理部が保持する管理情報と、前記分配比率決定部で決定された分配の比率とに基づいて、全記憶装置をまとめた場合の記憶容量を決定する容量計算部と、
前記帯域計算部で計算された値に基づいて、全記憶装置の入出力帯域の制御を行う全帯域制御部と、
前記容量計算部で計算された値に基づいて、全記憶装置の記憶容量の制御を行う全容量制御部とを備える、入出力帯域管理装置。

【請求項29】 前記分配比率決定部は、各前記記憶装置が有する帯域に比例して、ファイルデータを配置する比率を決定することを特徴とする、請求項28に記載の入出力帯域管理装置。

【請求項30】 前記分配比率決定部は、各前記記憶装置が有する記憶容量に比例して、ファイルデータを配置する比率を決定することを特徴とする、請求項28に記載

載の入出力帯域管理装置。

【請求項31】 前記分配比率決定部は、各前記記憶装置が有する帯域および記憶容量に比例して、ファイルデータを配置する比率を決定することを特徴とする、請求項28に記載の入出力帯域管理装置。

【請求項32】 前記帯域計算部は、各前記記憶装置が有する帯域に、前記分配比率決定部で決定された分配比率の逆数を掛けた値の中で、最も小さい値を、全記憶装置をまとめた場合の帯域とすることを特徴とする、請求項28に記載の入出力帯域管理装置。

【請求項33】 前記容量計算部は、各前記記憶装置が有する記憶容量に、前記分配比率決定部で決定された分配比率の逆数を掛けた値の中で、最も小さい値を、全記憶装置をまとめた場合の記憶容量とすることを特徴とする、請求項28に記載の入出力帯域管理装置。

【請求項34】 複数の記憶装置に対してファイルデータの入出力を行うファイルシステムであって、ファイルの属性として帯域特性を定義して管理するファイル属性管理部と、ファイルデータの入出力を行う前に前記ファイル属性管理部で管理されているファイルの帯域特性に従って、必要な帯域を獲得する帯域獲得部とを備える、ファイルシステム。

【請求項35】 前記ファイルの属性の帯域特性として、ファイルデータの各記憶装置への分配比率を記述したレイアウトプロファイル情報を用いることを特徴とする、請求項34に記載のファイルシステム。

【請求項36】 複数の記憶装置に対してファイルデータの入出力を行うファイルシステムであって、ファイルのオープン処理を行うファイルオープン処理部と、ファイルデータの入出力に使用する帯域の獲得を行う帯域獲得部と、前記ファイルオープン処理部によってオープンされた複数のファイルに対し、前記帯域獲得部で獲得した帯域を分割して個別的に割り当てることにより、複数のファイルのデータの入出力を同時に行うデータ入出力部とを備える、ファイルシステム。

【請求項37】 複数の記憶装置に対してファイルデータの入出力を行うファイルシステムであって、ファイルのオープン処理を行うファイルオープン処理部と、ファイルデータの入出力に使用する帯域の獲得を行う帯域獲得部と、ファイルオープン処理部によってオープンされた複数のファイルに対し、前記帯域獲得部で獲得した1つの帯域を統合的に割り当てることにより、同一の帯域を用いて複数のファイルのデータの入出力を交互に行うデータ入出力部とを備える、記憶装置システム。

【請求項38】 サーバから、複数のクライアントに、

それぞれ周期的にサービスが提供されるよう、サービスリクエストのスケジューリングを行うスケジューリング装置であって、

前記複数のクライアントからのサービスリクエストをキューイングする第1のキュー手段と、

前記複数のクライアント毎の、実行権を発生させるべき周期を登録する登録手段と、

前記登録手段に登録された周期に関連して、サービスリクエストを実行するための実行権を発生させる実行権発生手段と、

前記第1のキュー手段にキューイングされているサービスリクエストのうち、前記実行権発生手段が発生させた実行権に対応するクライアントからのサービスリクエストを、前記サーバに指示して実行させる実行指示手段とを備えるスケジューリング装置。

【請求項39】 前記実行権発生手段は、所定の周期で起動され、起動時に、前記登録手段に登録された、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、実行権を発生させるか否かをクライアント毎に判定し、かつ、実行権を発生させると判定したクライアントの実行権を、一括して発生させることを特徴とする、請求項38に記載のスケジューリング装置。

【請求項40】 前記実行権発生手段は、前記登録手段に登録された、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、起動された現時刻から、次の実行権を発生させる時刻までの時間を算定し、算定して得た時間と、自己が起動される周期とを比較することにより、実行権を発生させるか否かを判定することとする、請求項39に記載のスケジューリング装置。

【請求項41】 前記実行権発生手段は、起動された現時刻から、次の実行権を発生させる時刻までの時間 T_{ofs} が、自己が起動される周期 T_{slt} に対して、 $T_{ofs} < T_{slt}$ であるクライアントの実行権を発生させ、さらに、当該クライアントの T_{ofs} に、前記登録手段に登録された、実行権を発生させるべき周期 T_{req} を加算し、加算して得た T_{ofs} が、 $T_{ofs} \geq T_{slt}$ である場合、当該 T_{ofs} に対して、式 $T_{ofs} = T_{ofs} - T_{slt}$ により、新たな T_{ofs} を算定して、次回起動される際の T_{ofs} とすることを特徴とする、請求項40に記載のスケジューリング装置。

【請求項42】 前記実行権発生手段は、加算して得た T_{ofs} が、 $T_{ofs} < T_{slt}$ である場合には、さらに、実行権を発生させ、かつ当該 T_{ofs} に T_{req} を加算して T_{slt} と比較する動作を繰り返し、加算して得た T_{ofs} が、 $T_{ofs} \geq T_{slt}$ となった時点で、当該 T_{ofs} に対して、式 $T_{ofs} = T_{ofs} - T_{slt}$ により、新たな T_{ofs} を算定して、次回起動される際の T_{ofs} とすることを特徴とする、請求項41に記載のスケジューリング装置。

【請求項43】 前記実行権発生手段は、起動された現時刻から、次の実行権を発生させる時刻までの時間 $Tofs$ が、 $Tofs \geq Tslit$ であるか、または、サービスリクエストが、前記第1のキュー手段にキューイングされていないクライアントについては、実行権を発生させず、かつ、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tslit$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とすることを特徴とする、請求項41に記載のスケジューリング装置。

【請求項44】 前記実行権発生手段は、算定して得た新たな $Tofs$ が、 $Tofs < 0$ である場合には、次回起動される際の $Tofs$ を0とすることを特徴とする、請求項43に記載のスケジューリング装置。

【請求項45】 前記実行指示手段は、所定の周期で起動され、起動された現時刻から、次の実行権が発生される時刻までの時間が、より短いクライアントの実行権を選択することを特徴とする、請求項39に記載のスケジューリング装置。

【請求項46】 前記実行権発生手段は、優先度を示す評価指数を、前記登録手段に登録された、実行権を発生させるべき周期の関数として算定して、発生させた実行権に付与し、

前記実行指示手段は、付与された評価指数に基づいて、前記実行権発生手段が発生させた実行権のうち、優先されるべき実行権を選択することを特徴とする、請求項45に記載のスケジューリング装置。

【請求項47】 前記実行権発生手段は、評価指数 $Nsft$ を、前記登録手段に登録された実行権を発生させるべき周期 $Treq$ と、自己が起動される周期 $Tslit$ とに対して、関数 $Nsft = f(Treq/Tslit)$

(ただし、 $f(x)$ は x 以上の最小の整数) により算出し、

前記実行指示手段は、付与された評価指数 $Nsft$ がより小さい実行権を選択することを特徴とする、請求項46に記載のスケジューリング装置。

【請求項48】 前記実行権発生手段は、前記登録手段に登録された周期のうち、最も長いものよりは短い周期で、起動されることを特徴とする、請求項46に記載のスケジューリング装置。

【請求項49】 前記実行権発生手段は、前記登録手段に登録された周期のうち、最も短いものよりさらに短い周期で、起動されることを特徴とする、請求項47に記載のスケジューリング装置。

【請求項50】 前記実行指示手段は、前記実行権発生手段が発生させた実行権を、付与された評価指数毎で、かつ発生された順に、キューイングする第2のキュー手段を含み、当該キュー手段にキューイングされている実行権のうち、付与された評価指数がより小さく、かつ、より先に発生されたものを選択することを特徴とする、請求項46に記載のスケジューリング装置。

【請求項51】 前記実行指示手段は、起動時に実行を指示するサービスリクエストの最大数を認識しており、当該最大数に等しい個数だけサービスリクエストの実行を指示した時点で、実行の指示を停止し、かつ、残された実行権の評価指数を、より優先度の高いものに変更することを特徴とする、請求項46に記載のスケジューリング装置。

【請求項52】 前記実行指示手段が認識している、起動時に実行を指示するサービスリクエストの最大数は、自己が起動される周期に等しい時間内に前記サーバが実行可能なサービスリクエストの最大数であることを特徴とする、請求項51に記載のスケジューリング装置。

【請求項53】 前記実行指示手段は、起動時に実行を指示するサービスリクエストの最大数を認識しており、当該最大数に等しい個数だけサービスリクエストの実行を指示した時点で、実行の指示を停止し、かつ、残された実行権の評価指数をディクリメントすることを特徴とする、請求項47に記載のスケジューリング装置。

【請求項54】 前記実行指示手段が認識している、起動時に実行を指示するサービスリクエストの最大数は、自己が起動される周期に等しい時間内に前記サーバが実行可能なサービスリクエストの最大数であることを特徴とする、請求項53に記載のスケジューリング装置。

【請求項55】 前記複数のクライアントからの申告に基づいて、実行権を発生させるべき周期を、前記登録手段に指示して登録させる登録指示手段をさらに備えることを特徴とする、請求項51に記載のスケジューリング装置。

【請求項56】 前記登録指示手段は、単位時間当たり前に前記実行指示手段が実行を指示するサービスリクエストの最大数を認識しており、すでに登録されている周期に加えて、さらに他の周期を登録するよう申告された際、さらに登録されても、発生される実行権の単位時間当たりの平均数が、当該最大数を越えないと判定した場合、前記登録手段に指示して当該周期を登録させ、さらに登録されれば、越えると判定した場合には、登録されないことを特徴とする、請求項55に記載のスケジューリング装置。

【請求項57】 複数のサーバから、クライアントに、それぞれ周期的にサービスが提供されるよう、サービスリクエストのスケジューリングを行うスケジューリング装置であって、

前記複数のサーバへのサービスリクエストをキューイングする第1のキュー手段と、

前記複数のサーバ毎の、実行権を発生させるべき周期に登録する登録手段と、前記登録手段に登録された周期に関連して、サービスリクエストを実行するための実行権を発生させる実行権発生手段と、

前記第1のキュー手段にキューイングされているサービスリクエストのうち、前記実行権発生手段が発生させた

実行権に対応するサーバへのサービスリクエストを、当該サーバに指示して実行させる実行指示手段とを備えるスケジューリング装置。

【請求項58】 前記実行権発生手段は、所定の周期で起動され、起動時に、前記登録手段に登録された、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、実行権を発生させるか否かをサーバ毎に判定し、かつ、実行権を発生させると判定したサーバの実行権を、一括して発生させることを特徴とする、請求項57に記載のスケジューリング装置。

【請求項59】 前記実行権発生手段は、前記登録手段に登録された、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、起動された現時刻から、次の実行権を発生させる時刻までの時間を算定し、算定して得た時間と、自己が起動される周期とを比較することにより、実行権を発生させるか否かを判定することを特徴とする、請求項58に記載のスケジューリング装置。

【請求項60】 前記実行権発生手段は、起動された現時刻から、次の実行権を発生させる時刻までの時間 T_{ofs} が、自己が起動される周期 T_{slt} に対して、 $T_{ofs} < T_{slt}$ であるサーバの実行権を発生させ、さらに、当該サーバの T_{ofs} に、前記登録手段に登録された、実行権を発生させるべき周期 T_{itv} を加算し、加算して得た T_{ofs} が、 $T_{ofs} \geq T_{slt}$ である場合、当該 T_{ofs} に対して、式 $T_{ofs} = T_{ofs} - T_{slt}$ により、新たな T_{ofs} を算定して、次回起動される際の T_{ofs} とすることを特徴とする、請求項59に記載のスケジューリング装置。

【請求項61】 前記実行権発生手段は、加算して得た T_{ofs} が、 $T_{ofs} < T_{slt}$ である場合には、さらに、実行権を発生させ、かつ当該 T_{ofs} に T_{itv} を加算して T_{slt} と比較する動作を繰り返し、加算して得た T_{ofs} が、 $T_{ofs} \geq T_{slt}$ となった時点で、当該 T_{ofs} に対して、式 $T_{ofs} = T_{ofs} - T_{slt}$ により、新たな T_{ofs} を算定して、次回起動される際の T_{ofs} とすることを特徴とする、請求項60に記載のスケジューリング装置。

【請求項62】 前記実行権発生手段は、起動された現時刻から、次の実行権を発生させる時刻までの時間 T_{ofs} が、 $T_{ofs} \geq T_{slt}$ であるか、または、サービスリクエストが、前記第1のキュー手段にキューイングされていないサーバについては、実行権を発生させず、かつ、当該 T_{ofs} に対して、式 $T_{ofs} = T_{ofs} - T_{slt}$ により、新たな T_{ofs} を算定して、次回起動される際の T_{ofs} とすることを特徴とする、請求項60に記載のスケジューリング装置。

【請求項63】 前記実行権発生手段は、算定して得た新たな T_{ofs} が、 $T_{ofs} < 0$ である場合には、次回起動される際の T_{ofs} を0とすることを特徴とする、

請求項62に記載のスケジューリング装置。

【請求項64】 前記実行指示手段は、所定の周期で起動され、起動された現時刻から、次の実行権が発生される時刻までの時間が、より短いサーバの実行権を選択することを特徴とする、請求項58に記載のスケジューリング装置。

【請求項65】 前記実行権発生手段は、優先度を示す評価指数を、前記登録手段に登録された、実行権を発生させるべき周期の関数として算定して、発生させた実行権に付与し、

前記実行指示手段は、付与された評価指数に基づいて、前記実行権発生手段が発生させた実行権のうち、優先されるべき実行権を選択することを特徴とする、請求項64に記載のスケジューリング装置。

【請求項66】 前記実行権発生手段は、評価指数 N_{sft} を、前記登録手段に登録された実行権を発生させるべき周期 T_{itv} と、自己が起動される周期 T_{slt} とに対して、関数 $N_{sft} = f(T_{itv}/T_{slt})$

(ただし、 $f(x)$ は x 以上の最小の整数) により算出し、

前記実行指示手段は、付与された評価指数 N_{sft} がより小さい実行権を選択することを特徴とする、請求項65に記載のスケジューリング装置。

【請求項67】 前記実行権発生手段は、前記登録手段に登録された周期のうち、最も長いものよりは短い周期で、起動されることを特徴とする、請求項66に記載のスケジューリング装置。

【請求項68】 前記実行権発生手段は、前記登録手段に登録された周期のうち、最も短いものよりさらに短い周期で、起動されることを特徴とする、請求項66に記載のスケジューリング装置。

【請求項69】 前記実行指示手段は、前記実行権発生手段が発生させた実行権を、付与された評価指数毎で、かつ発生された順に、キューイングする第2のキュー手段を含み、当該キュー手段にキューイングされている実行権のうち、付与された評価指数がより小さく、かつ、より先に発生されたものを選択することを特徴とする、請求項65に記載のスケジューリング装置。

【請求項70】 前記実行指示手段は、起動時に実行を指示するサービスリクエストの最大数を認識しており、当該最大数に等しい個数だけサービスリクエストの実行を指示した時点で、実行の指示を停止し、かつ、残された実行権の評価指数を、より優先度の高いものに変更することを特徴とする、請求項65に記載のスケジューリング装置。

【請求項71】 前記実行指示手段が認識している、起動時に実行を指示するサービスリクエストの最大数は、自己が起動される周期に等しい時間内に、全サーバからクライアントに送信可能な最大の情報量に対応するサービスリクエストの個数であることを特徴とする、請求項

70に記載のスケジューリング装置。

【請求項72】 前記実行指示手段は、起動時に実行を指示するサービスリクエストの最大数を認識しており、当該最大数に等しい個数だけサービスリクエストの実行を指示した時点で、実行の指示を停止し、かつ、残された実行権の評価指数をディクリメントすることを特徴とする、請求項66に記載のスケジューリング装置。

【請求項73】 前記実行指示手段が認識している、起動時に実行を指示するサービスリクエストの最大数は、自己が起動される周期に等しい時間内に、全サーバからクライアントに送信可能な最大の情報量に対応するサービスリクエストの個数であることを特徴とする、請求項72に記載のスケジューリング装置。

【請求項74】 前記クライアントからの申告に基づいて、実行権を発生させるべき周期を、前記登録手段に指示して登録／更新させる登録指示手段をさらに備えることを特徴とする、請求項70に記載のスケジューリング装置。

【請求項75】 前記登録指示手段は、単位時間当たりに前記実行指示手段が全サーバに実行を指示するサービスリクエストの最大数を認識しており、すでに登録されている周期に加えて、さらに他のサーバの周期を登録するよう申告された際、さらに登録されても、発生される、全サーバの実行権の単位時間当たりの平均数が、当該最大数を越えないと判定した場合、前記登録手段に指示して当該周期を登録させ、さらに登録されれば、越えると判定した場合には、登録させないことを特徴とする、請求項74に記載のスケジューリング装置。

【請求項76】 前記登録指示手段は、単位時間当たりに前記実行指示手段が各サーバに実行を指示するサービスリクエストの最大数を認識しており、すでに登録されている周期に加えて、さらに他のサーバの周期を登録するよう申告された際、さらに登録されても、発生される、当該サーバの実行権の単位時間当たりの平均数が、単位時間当たりに前記実行指示手段が当該サーバに実行を指示するサービスリクエストの最大数を越えないと判定した場合、前記登録手段に指示して当該周期を登録させ、さらに登録されれば、越えると判定した場合には、登録させないことを特徴とする、請求項74に記載のスケジューリング装置。

【請求項77】 前記登録指示手段が認識している、単位時間当たりに前記実行指示手段が各サーバに実行を指示するサービスリクエストの最大数は、単位時間当たりに前記各サーバが実行可能なサービスリクエストの最大数であることを特徴とする、請求項76に記載のスケジューリング装置。

【請求項78】 前記登録指示手段は、単位時間当たりに前記実行指示手段が全サーバに実行を指示するサービスリクエストの最大数と、単位時間当たりに前記実行指示手段が各サーバに実行を指示するサービスリクエスト

の最大数とを認識しており、すで登録されている、あるサーバの実行権を発生させるべき周期 T_{itv} を、さらに周期 T_{req} を加えて実行権を発生させるものに更新するよう、申告を受けた際、更新されても、発生される、全サーバの実行権の単位時間当たりの平均数が、単位時間当たりに前記実行指示手段が前記全サーバに実行を指示するサービスリクエストの最大数を越えないと判定し、かつ、発生される、当該サーバの実行権の単位時間当たりの平均数が、単位時間当たりに前記実行指示手段が当該サーバに実行を指示するサービスリクエストの最大数を越えないと判定した場合、前記登録手段に指示して、すでに登録されている T_{itv} を、式 $T_{itv} = 1 / (1 / T_{itv} + 1 / T_{req})$ により算定される新たな T_{itv} に更新させ、更新されれば、越えると判定した場合には、更新させないことを特徴とする、請求項74に記載のスケジューリング装置。

【請求項79】 前記登録指示手段は、すで登録されている、あるサーバの実行権を発生させるべき周期 T_{itv} を、周期 T_{req} を除いて実行権を発生させるものに更新するよう申告を受けた際、前記登録手段に指示して、すでに登録されている T_{itv} を、式 $T_{itv} = 1 / (1 / T_{itv} - 1 / T_{req})$ により算定される新たな T_{itv} に更新させることを特徴とする、請求項74に記載のスケジューリング装置。

【請求項80】 有線または無線の伝送路を介して複数の端末と接続されており、いずれかの端末からデータの読み出し要求が送られてきたとき、内部に蓄積されたデータの中から対応するデータを検索して読み出し、読み出したデータを読み出し要求を発した端末へと送信するサーバ装置であって、

種々のデータを記憶しており、前記端末からの読み出し要求に応じて、ランダムアクセスが可能なデータ記憶部と、

前記データ記憶部と各前記端末との間でデータの入出力を制御する入出力部と、

前記入出力部と前記データ記憶部との間でデータを交換する交換部と、

前記端末から要求されたデータを、前記データ記憶部から読み出して当該端末へと出力する際に必要となるリソースを管理するリソース管理部とを備え、

前記リソース管理部は、サーバ装置全体のリソースを、連続性を保証する必要があるデータ（優先データ）を前記端末へと出力するために用いられるリソース（優先リソース）と、連続性を保証する必要のないデータ（非優先データ）を前記端末へと出力するために用いられるリソース（非優先リソース）とに分離して管理することを特徴とする、サーバ装置。

【請求項81】 各前記端末は、複数のグループに分割されており、

前記入出力部は、前記端末のグループ毎に設けられた複

数のI/Oユニットを含み、
 各前記I/Oユニットは、
 対応するグループの端末から送られてくる優先データの読み出しを要求するコマンドに対する入出力制御を実行する優先入出力制御部と、
 対応するグループの端末から送られてくる非優先データの読み出しを要求するコマンドに対する入出力制御を実行する非優先入出力制御部とを含み、
 前記データ記憶部は、
 複数の記憶装置と、
 各前記記憶装置と接続され、当該記憶装置からのデータの読み出しを制御する複数のMSFSとを含み、
 前記非優先入出力制御部は、
 対応するグループの端末から送られてくる非優先データの読み出しを要求するコマンドを一時的に格納するコマンドキューと、
 前記コマンドキューからコマンドを取り出すコマンドキュー制御部と、
 前記コマンドキュー制御部が取り出したコマンドを、対応する宛先のMSFSに送出するコマンド送出部と、
 前記MSFSによって前記記憶装置から読み出された非優先データを一時的に格納するデータバッファ部と、
 前記データバッファ部に格納された非優先データを前記端末に送出するデータ送出部とを含む、請求項80に記載のサーバ装置。

【請求項82】 前記コマンドキューは、対応するグループに属する複数の端末のそれぞれに対応して設けられており、それぞれが対応する端末からのコマンドを到着順に格納し、
 前記コマンドキュー制御部は、複数のコマンドキューの中からコマンドを格納している1つのコマンドキューを選択すると共に、その先頭に格納されているコマンドおよびそれと同じMSFSに送出されるコマンドを取り出して前記コマンド送出部に出力するコマンド取り出し処理を実行し、
 前記コマンド送出部は、
 前記コマンドキュー制御部から受け取ったコマンドを、それを実行すべきMSFSに送出し、
 送出したコマンドに対する全ての読み出しデータを、前記データバッファ部が前記MSFSから受信したことを確認すると、受信完了通知を前記コマンドキュー制御部に送出し、
 前記コマンドキュー制御部は、前記コマンド送出部から前記受信完了通知を受け取る毎に、前記コマンド取り出し処理を繰り返し実行することを特徴とする、請求項81に記載のサーバ装置。

【請求項83】 前記コマンドキュー制御部は、1回の前記コマンド取り出し処理において、1つのコマンドキューから取り出すコマンドの数を、予め定めた一定個数以下に制限することを特徴とする、請求項82に記載の

サーバ装置。

【請求項84】 前記非優先入出力制御部は、前記データバッファ部の使用状態を管理するバッファ管理部をさらに含み、
 前記コマンドキュー制御部は、前記バッファ管理部が管理する前記データバッファ部の使用状態に基づき、前記コマンド取りだし処理のために選択するコマンドキューを決定することを特徴とする、請求項82に記載のサーバ装置。

【請求項85】 前記バッファ管理部には、各端末毎に前記データバッファ部に格納されている未送信データ数をカウントする複数の使用量カウンタが設けられており、
 前記コマンドキュー制御部は、前記コマンド取り出し処理を実行する前に前記バッファ管理部の使用量カウンタを参照し、未送信データ数が最も小さい端末に対応するコマンドキューを、コマンド取り出し処理の対象として選択することを特徴とする、請求項84に記載のサーバ装置。

【請求項86】 前記コマンド送出部は、前記MSFSによって読み出される非優先データを格納するための領域を前記データバッファ部に予約した後に、前記コマンドキュー制御部から受け取ったコマンドを前記MSFSに送出し、データバッファ部に非優先データを格納するための領域が予約できない場合には、予約できるまでMSFSに対するコマンドの送出を停止することを特徴とする、請求項82に記載のサーバ装置。

【請求項87】 前記コマンドキュー制御部は、前記I/Oユニットが前記端末にデータを出力するための非優先リソースとしての通信帯域と、前記I/Oユニットが前記データ記憶部から読み出された非優先データを受信するための非優先リソースとしての通信帯域との内、小さい方の帯域に基づいて、前記コマンド送出部に対してコマンドを出力する周期を決定することを特徴とする、請求項81に記載のサーバ装置。

【請求項88】 前記コマンド送出部は、
 各前記MSFSに対応して設けられ、前記コマンドキュー制御部から受け取ったコマンドを、MSFS別に格納する複数の送出キューと、
 各前記送出キューのそれぞれに対応して設けられ、対応する送出キューからMSFSへコマンドを送出する毎にインクリメントされ、対応するMSFSからの読み出しデータが前記データバッファ部に到着する毎にデクリメントされることにより、データバッファ部への非優先データの未到着数をカウントする複数の処理数カウンタとを含み、
 前記コマンド送出部は、前記処理数カウンタによってカウントされる未到着数が所定のしきい値を超えないように、各送出キューからMSFSに対して送出されるコマンドの数を制限することを特徴とする、請求項81に記載

載のサーバ装置。

【請求項 89】 前記リソース管理部は、前記リソースとして、前記交換部の通信帯域と、前記データ記憶部がデータを読み出す帯域と、前記入出力部が各前記端末にデータを出力するための通信帯域とを管理していることを特徴とする、請求項 80 に記載のサーバ装置。

【請求項 90】 前記交換部は、ATMスイッチで構成されており、

前記 ATM スwitch の非優先リソースとしての通信帯域が、ABR モードで確保されることを特徴とする、請求項 80 に記載のサーバ装置。

【請求項 91】 前記リソース管理部は、予め固定的に、サーバ装置全体のリソースの内、一部を前記優先リソースとして割り当て、残りの部分を前記非優先リソースとして割り当てることを特徴とする、請求項 80 に記載のサーバ装置。

【請求項 92】 前記リソース管理部は、前記端末からの前記優先データの読み出し要求に応じて、サーバ装置全体のリソースの内、前記優先リソースとして割り当てる量を変化させることを特徴とする、請求項 80 に記載のサーバ装置。

【請求項 93】 前記リソース管理部は、サーバ装置全体のリソースの内、前記優先リソースとして割り当てられていないリソースを全て前記非優先リソースとして割り当て、前記優先リソースを増やす必要が生じた場合、前記非優先リソースが所定のリソース量以下にならない範囲で、当該非優先リソースとして割り当てられているリソースから優先リソースを割り当てることを特徴とする、請求項 92 に記載のサーバ装置。

【請求項 94】 利用者側からのデータ送出要求に回答してデータ取得要求をデータ記憶装置に発行し、当該データ取得要求に従ってデータ記憶装置から取得したデータを利用者側に送出するファイルシステムであって、前記データ送出要求によって指定された位置から順番に、かつ指定された速度で連続的に前記データ記憶装置からデータを取得するデータ取得部と、前記データ取得部がデータ記憶装置から取得したデータを格納するバッファと、前記バッファに格納されたデータの操作とデータ量の管理とを行うバッファ管理部と、前記バッファに格納されているデータを、前記データ送出要求によって指定された順番に、かつ指定された伝送速度で連続的に読み出して利用者側に送出するデータ送出部と、利用者側からのポーズ要求およびポーズ解除要求を受けて、前記データ送出部、前記データ取得部および前記バッファ管理部を制御するポーズ制御部とを備え、前記ポーズ制御部は、利用者側からのポーズ要求を受けた場合、直ちにデータ送出を停止するよう前記データ送出部を制御すると共

に、直ちにデータ取得を停止するよう前記データ取得部を制御し、

利用者側からのポーズ解除要求を受けた場合、ポーズ処理以降に前記バッファに格納されているデータの先頭から順番にデータ送出を再開するよう前記データ送出部を制御すると共に、ポーズ処理以降にバッファに格納されているデータの最後の次のデータから順番にデータ取得を再開するよう前記データ取得部を制御することを特徴とする、ファイルシステム。

【請求項 95】 前記ポーズ制御部は、利用者側からのポーズ解除要求を受けた場合、前記データ取得部にデータ取得を再開させるデータ取得再開時刻 t_{rd} と、ポーズ処理以降に前記バッファに格納されているデータ量 N と、ポーズ解除後のデータ伝送速度 V と、データ取得部がデータ取得に必要なとする時間 D とを、次式 (a) に代入することにより、前記データ送出部にデータ送出を再開させるデータ送出再開時刻 t_{sd} を算出することを特徴とする、請求項 94 に記載のファイルシステム。

$$t_{sd} = t_{rd} + D - N/V \quad \dots (a)$$

【請求項 96】 前記ポーズ制御部は、利用者側からのポーズ解除要求を受けた場合、前記データ送出部にデータ送出を再開させるデータ送出再開時刻 t_{sd} と、ポーズ処理以降に前記バッファに格納されているデータ量 N と、ポーズ解除後のデータ伝送速度 V と、前記データ取得部がデータ取得に必要なとする時間 D とを、次式 (b) に代入することにより、データ取得部にデータ取得を再開させるデータ取得再開時刻 t_{rd} を算出することを特徴とする、請求項 94 に記載のファイルシステム。

$$t_{rd} = t_{sd} + N/V - D \quad \dots (b)$$

【請求項 97】 前記ポーズ制御部は、利用者側からのポーズ解除要求を受けた場合、ポーズ処理以降に前記バッファに格納されているデータ量が、前記データ取得部がデータ取得に必要なとする時間中にポーズ解除後のデータ伝送速度で送出可能な限界データ量よりも大きいかな否かを判断し、前記判断の結果、前記データ量が前記限界データ量よりも大きい場合は、当該限界データ量を越える部分のデータをバッファから削除するように前記バッファ管理部を制御し、それによってポーズ解除後にデータ取得部が取得するデータの先頭位置を決定することを特徴とする、請求項 94 に記載のファイルシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、ファイルシステムに関し、より特定的には、リアルタイム性が要求される映像、音声、動画といったマルチメディアデータの読み書きを扱うファイルシステムに関する。

【0002】

【従来の技術】 従来のコンピュータで用いられているファイルシステムは、主としてテキストデータを扱ってき

た。例えば、「UNIXカーネルの設計」(Maurice J. Bach 著／坂本文・多田好克・村井純訳；共立出版株式会社)によると、UNIXのファイルシステムでは、ファイルからデータを読み出し、またはファイルにデータを書き込む場合には、openシステムコールを用いてファイルを開き、ファイルの内容を、read, writeといったシステムコールを用いて読み出しまたは書き込むようにしている。

【0003】ところで、ファイルシステムの一例として、デジタル化された映像データを扱うビデオサーバがある。このようなビデオサーバには、特開平07-107425号公報の「ビデオサーバ」に記載されているように、ビデオデータを格納した複数の記憶装置を、交換装置で接続した構成のものがある。また、ビデオサーバにおける映像の再生手順については、ISO/IEC標準13818-6(Committee Draft)において記述されている。このCommittee Draftによれば、ユーザがビデオサーバに「Directory Open」を送出すると、ビデオサーバは、映像の連続再生の準備を完了し、「DSM Stream Play」「DSM Pause Resume」を送出すると、ビデオサーバからユーザへの映像データの送出手続きが開始される。また、映像データの供給を停止させる場合には、ユーザからビデオサーバに「DSM Stream Pause」が送出される。

【0004】

【発明が解決しようとする課題】しかしながら、UNIXのような従来のファイルシステムでは、ユーザからのアクセスが多数同時に発生した場合には、ユーザ側での読み出し待ち時間が増加する。このように、ユーザが読み出しコマンドを発行してから、データの受け取りを完了するまでの待ち時間は、他のユーザからのアクセス状況に影響を受けるため、予測不可能であった。特に、ビデオサーバのように、ファイルシステムが映像や、音声といった時間的に連続性が要求されるデータを管理しており、ユーザがデータの一部を一定時間おきに読み出して、ユーザ端末で再生を行う場合には、一定時間内に確実にデータを取得する必要があるが、従来のファイルシステムでは対応できなかった。従って、映像データのように時間的に連続性が要求されるデータに関しては、ユーザ側で予め長いデータを取得しておき、全て取得した後に、映像の再生を行う必要があった。この場合には、ユーザ側で長いデータを格納できる大容量の記憶装置を保持する必要があるが、また、映像の再生を開始するまでに長い待ち時間を要した。

【0005】それ故に、本発明の目的は、ユーザからの要求に対して素早い応答時間でデータを提供でき、しかも様々な再生速度を要求するユーザをサポートすることのできるファイルシステムを提供することである。

【0006】

【課題を解決するための手段および発明の効果】第1の発明は、外部からの要求に応じて、データを記憶する記憶装置に対してデータの読み出しや書き込みを行うファイルシステムであって、読み出しや書き込みの処理能力の最大値と、当該処理能力のうちの使用量とを管理するリソース管理部を備え、リソース管理部は、データの読み出しや書き込みを行う際には、予め必要な帯域の確保を要求するゲットリソースメッセージを受信し、ゲットリソースメッセージを受信したとき、処理能力の最大値および使用量に基づいて、当該ゲットリソースメッセージで要求される帯域を確保できるかを判断し、判断の結果、要求される帯域を確保できる場合には、使用量を更新して記憶することを特徴とする。

【0007】上記のように、第1の発明によれば、ゲットリソースメッセージを受信したときに、ファイルシステム内部の帯域を確保するようにしているので、ファイルの読み出し、書き込みについて、遅延品質を保証することが可能となる。そして、確保した帯域を用いて複数のファイルの読み出しまたは書き込みを実現することにより、無駄な帯域を確保する必要がなく、ファイルシステム内部のリソースを有効利用することができる。

【0008】第2の発明は、第1の発明において、ゲットリソースメッセージには、パラメータとして、必要な帯域量を示す情報が付与されていることを特徴とする。

【0009】第3の発明は、第2の発明において、ゲットリソースメッセージのパラメータである、必要な帯域量を示す情報は、記憶装置に対して、固定長のデータブロックを読み出しまたは書き込みする間隔とすることを特徴とする。

【0010】第4の発明は、第1の発明において、ゲットリソースメッセージの要求に従って確保する帯域量は、リソース管理部において予め記憶されていることを特徴とする。

【0011】第5の発明は、第4の発明において、ゲットリソースメッセージには、読み出し用、書き込み用を識別する情報が付与されており、リソース管理部は、ゲットリソースメッセージの要求に従って確保する帯域量として、読み出し用の帯域量と書き込み用の帯域量との2種類を記憶していることを特徴とする。

【0012】第6の発明は、第1の発明において、ゲットリソースメッセージには、パラメータとして、読み出しの要求を行ってからデータの送信が開始されるまでの第1の遅延時間が付与されており、リソース管理部は、予めファイルシステムが保証可能な遅延時間の最短値を記憶しており、ゲットリソースメッセージを受信したとき、第1の遅延時間と最短値とを比較し、比較の結果、第1の遅延時間が最短値よりも短い場合、帯域の確保を行わないことを特徴とする。

【0013】第7の発明は、第6の発明において、外部から到着するメッセージに付与されているパラメータを

チェックする申告値管理部をさらに備え、申告値管理部は、データの読み出しを行う際には、読み出されたデータの送信が開始されるまでの第2の遅延時間がパラメータとして付与されたリードブロックメッセージを受信し、リードブロックメッセージを受信したとき、第2の遅延時間と第1の遅延時間とを比較し、比較の結果、第2の遅延時間のほうが短いと判断した場合、リードブロックメッセージの送信元に対して違反通知を送信することを決定することを特徴とする。

【0014】第8の発明は、第1の発明において、外部から到着するメッセージに付与されているパラメータをチェックする申告値管理部をさらに備え、申告値管理部は、ファイルシステムが保証可能な遅延時間の最短値を記憶しており、データの読み出しを行う際には、読み出されたデータの送信が開始されるまでの第2の遅延時間がパラメータとして付与されたリードブロックメッセージを受信し、リードブロックメッセージを受信したとき、第2の遅延時間と最短値とを比較し、比較の結果、第2の遅延時間のほうが短いと判断した場合、リードブロックメッセージの送信元に対して違反通知を送信することを決定することを特徴とする。

【0015】第9の発明は、第1の発明において、ゲットリソースメッセージには、パラメータとして、読み出しの要求を行ってからデータの送信が開始されるまでの遅延時間が付与されており、リソース管理部は、ゲットリソースメッセージを受信したとき、遅延時間を満たすために必要な帯域量を算出し、算出の結果得た帯域量とゲットリソースメッセージの要求に従って確保すべき帯域量とを比較して、大きい方の帯域量を確保できるかどうかを判断することを特徴とする。

【0016】上記のように、第9の発明によれば、ゲットリソースメッセージにより申告された遅延時間を満たす帯域と、ゲットリソースメッセージの要求に従って確保すべき帯域との内、大きい方の帯域を確保できるかどうかを判断して、帯域の割り当てを行うようにしているので、複数の遅延品質をサポートすることが可能となる。

【0017】第10の発明は、第1の発明において、外部から到着するゲットリソースメッセージに付与されているパラメータをチェックする申告値管理部をさらに備え、申告値管理部は、データの書き込み時において、データ到着の速度を観測し、観測の結果得たデータ到着の速度と、ゲットリソースメッセージ受付時に確保した帯域量とを比較し、比較の結果、データ到着の速度のほうが大きいと判断したときは、データ書き込みは行わず、ゲットリソースメッセージ送信元に対して違反通知を行うことを特徴とする。

【0018】第11の発明は、第1の発明において、外部から到着するメッセージに付与されているパラメータをチェックする申告値管理部をさらに備え、申告値管理部は、データの読み出し時には、単位時間当たりの読み

出しデータ量がパラメータとして付与されたリードブロックメッセージを受信し、リードブロックメッセージを受信したとき、単位時間当たりのデータ量からデータ読み出し速度を算出し、算出によって得たデータ読み出し速度と、ゲットリソースメッセージ受付時に確保した帯域量とを比較し、比較の結果、データ読み出し速度のほうが大きい場合、データの読み出しを中止し、リードブロックメッセージ送信元に対して違反通知を送信することを決定することを特徴とする。

【0019】第12の発明は、第1の発明において、処理能力の最大値および使用量は、単位時間当たりの読み出しデータ量であることを特徴とする。

【0020】第13の発明は、第12の発明において、ゲットリソースメッセージには、パラメータとして、読み出し、書き込みを区別するフラグと、必要な帯域量を示す情報とが付与されており、リソース管理部は、ゲットリソースメッセージを受信したとき、フラグを参照し、読み出しの場合には、ゲットリソースメッセージに付与されているパラメータによって決定される帯域量を確保すべき帯域量とし、書き込みの場合には、ゲットリソースメッセージに付与されているパラメータによって決定される帯域量に一定値を乗じたものを確保すべき帯域量とすることを特徴とする。

【0021】第14の発明は、第1の発明において、処理能力の最大値および使用量は、単位時間当たりの書き込みデータ量であることを特徴とする。

【0022】第15の発明は、第14の発明において、ゲットリソースメッセージには、パラメータとして、読み出し、書き込みを区別するフラグと、必要な帯域量を示す情報とが付与されており、リソース管理部は、ゲットリソースメッセージを受信したとき、フラグを参照し、書き込みの場合には、ゲットリソースメッセージに付与されているパラメータによって決定される帯域量を確保すべき帯域量とし、読み出しの場合には、ゲットリソースメッセージに付与されているパラメータによって決定される帯域量に一定値を乗じたものを確保すべき帯域量とすることを特徴とする。

【0023】上記のように、第13または第15の発明によれば、データの書き込み時または読み出し時には、要求帯域に一定値を乗じたものを確保すべき帯域とするようにしているので、予め書き込み用と、読み出し用とにファイルシステムの処理能力を分割しておく必要がなく、ユーザからの要求に応じて容易に比率を変えることができ、リソースを効率よく使用することができる。

【0024】第16の発明は、データを格納した複数のブロック格納ユニットと、ユーザからの要求に応じてデータの書き込みや読み出しを制御する単数または複数のI/Oユニットと、各ブロック格納ユニットおよび各I/Oユニットの間を接続する交換装置とを備え、各I/Oユニットは、ユーザから要求されたデータの大きさを

固定長のブロックに分割し、ブロック毎にブロック格納ユニットに対してブロック読み出し要求を送出し、ブロックの読み出しを行うファイルシステムにおいて、I/Oユニットは、一定時間毎に時間を区切ったMS到着スロットを設定し、当該MS到着スロットには、一定数のブロック読み出し要求を登録できるものとし、ブロック読み出し要求送出の際には、当該ブロック読み出し要求に対してMS到着スロットを割り当て、割り当てたMS到着スロットに対応した時刻に基づき、ブロック格納ユニットから交換装置への入力が許可される時刻を算出し、算出された時刻と現在時刻との差の時間を遅延限界値としてブロック読み出し要求に付与してブロック格納ユニット宛に送出し、それによってブロック読み出し要求に応答してブロック格納ユニットから読み出されるブロックは、当該ブロック読み出し要求が当該ブロック格納ユニットに到着してから遅延限界値の時間の後に、交換装置に対する入力が許可されることを特徴とする。

【0025】上記のように、第16の発明によれば、ブロック読み出し要求に対してMS到着スロットを割り当てるようにしているので、ブロック格納ユニットから交換装置に入力されるブロックの個数を時間的に平滑化することができる。その結果、複数のブロック格納ユニットから読み出されて交換装置に入力されるブロックが、I/Oユニットへの入力側でぶつかって交換装置内部で廃棄されることを回避することができる。また、ブロックのトラヒックの時間による変動を少なくすることができ、交換装置を効率よく利用することが可能となる。

【0026】第17の発明は、第16の発明において、その内部にI/Oユニットおよびブロック格納ユニットが設けられたサーバユニットを複数組備えている。

【0027】第18の発明は、第16の発明において、I/Oユニットは、ブロック読み出し要求の内、交換装置に入力されないブロック読み出し要求に対しては、MS到着スロットの割り当てを行わないことを特徴とする。

【0028】第19の発明は、第16の発明において、読み出しや書き込みの処理能力の最大値と、当該処理能力のうちの使用量とを管理するリソース管理部をさらに備え、リソース管理部は、データの読み出しや書き込みを行う際には、予め必要な帯域の確保を要求するゲットリソースメッセージを受信し、ゲットリソースメッセージを受信したとき、処理能力の最大値および使用量に基づいて、当該ゲットリソースメッセージで要求される帯域を確保できるかを判断し、判断の結果、要求される帯域を確保できる場合には、使用量を更新して記憶し、I/Oユニットは、ファイルの読み出しに必要な帯域を確保している場合には、確保した帯域だけのデータを読み出すために必要となるブロック読み出し要求の発行間隔をT_{min}に設定し、当該T_{min}の間隔で、MS到着スロットの期待スロットを決定し、決定した期待ス

ロットに空きが存在すれば、当該期待スロットを予約し、決定した期待スロットに空きが存在しない場合には、順次、次の時刻のMS到着スロットを参照して、空きが存在するMS到着スロットを予約し、データ読み出しの際には、予約されたMS到着スロットにブロック読み出し要求の割り当てを行うことを特徴とする。

【0029】上記のように、第19の発明によれば、ゲットリソースメッセージを受信したときに、ファイルシステム内部の帯域を確保するようにしているので、ファイルの読み出し、書き込みについて、遅延品質を保証することが可能となる。そして、確保した帯域を用いて複数のファイルの読み出しまたは書き込みを実現することにより、無駄な帯域を確保する必要がなく、ファイルシステム内部のリソースを有効利用することができる。

【0030】第20の発明は、第19の発明において、I/Oユニットは、MS到着スロットの予約を、確保された帯域が大きいものから小さいものの順に行うことを特徴とする。

【0031】上記のように、第20の発明によれば、MS到着スロットの予約を、確保された帯域が大きいものから小さいものの順で行うようにしているので、すべての帯域の遅延品質を保証しつつ、各スロットの最大登録数を一定にして、スロットを100%利用することが可能となる。このとき、要求送出数の時間変動はなくなり、最も効率良く交換装置を利用することが可能となる。

【0032】第21の発明は、第19の発明において、I/Oユニットは、MS到着スロットの予約を、次の期待スロットが到来するタイミングが早いものから順番に行うことを特徴とする。

【0033】第22の発明は、第19の発明において、I/Oユニットは、MS到着スロットの割り当てを行った結果、割り当てたMS到着スロットよりも前の時刻で、かつ同一ユーザのために予約されたMS到着スロットが存在する場合には、当該予約されたMS到着スロットを割り当てがされなかったMS到着スロットとし、リソースの獲得を行っていないユーザからのデータの読み出し要求に対し、当該割り当てがされなかったMS到着スロットを割り当てることを特徴とする。

【0034】第23の発明は、第19の発明において、I/Oユニットは、MS到着スロットの予約をゲットリソースメッセージを受け付けているユーザすべてについて行った結果、予約が行われなかったMS到着スロットに対しては、リソースの獲得を行っていないユーザからのデータの読み出し要求に割り当てることを特徴とする。

【0035】上記のように、第22または第23の発明によれば、使用しないMS到着スロットを、リソースの獲得を行っていないユーザからのデータの読み出しに割り当てるようにしているので、さらにファイルシステム

の使用効率を向上させることが可能となる。

【0036】第24の発明は、第1の発明において、データを格納した単数または複数のブロック格納ユニットと、ユーザからの要求に応じてデータの書き込みや読み出しを制御する単数または複数のI/Oユニットと、各ブロック格納ユニットおよび各I/Oユニットの間を接続する交換装置とをさらに備え、I/Oユニットは、ユーザから書き込みを要求されたデータの大きさを固定長のブロックに分割し、当該ブロック毎にブロック格納ユニットに対してブロック読み出し要求を送出することにより、ブロックの読み出しを行い、ブロック格納ユニットに対応させて、一定時間毎に区切られた書き込み用スロットを設定し、ブロック書き込み要求を送出する際には、当該書き込み用スロットを割り当てて送出的ことを特徴とする。

【0037】上記のように、第24の発明によれば、ブロック読み出し要求に対してMS到着スロットを割り当てるようにしているので、ブロック格納ユニットから交換装置に入力されるブロックの個数を時間的に平滑化することができる。その結果、複数のブロック格納ユニットから読み出されて交換装置に入力されるブロックが、I/Oユニットへの入力側でぶつかって交換装置内部で廃棄されることを回避することができる。また、ブロックのトラヒックの時間による変動を少なくすることができ、交換装置を効率よく利用することが可能となる。

【0038】第25の発明は、複数の記憶装置に対してファイルデータの入出力を行うファイルシステムにおいて、ファイルデータを各記憶装置に分散配置する方法であって、各記憶装置が有する帯域に比例してファイルデータを配置する比率を決定し、当該決定された比率に従ってファイルデータを各記憶装置に分散して格納することを特徴とする。

【0039】上記のように、第25の発明によれば、各記憶装置が有する帯域に比例してファイルデータを配置する比率を決定し、当該決定された比率に従ってファイルデータを各記憶装置に分散して格納するようにしているので、性能（データの読出／書込速度）の低い記憶装置がボトルネックとなって性能の高い記憶装置の能力を十分に引き出せなくなるといった事態を防止することができる。

【0040】第26の発明は、複数の記憶装置に対してファイルデータの入出力を行うファイルシステムにおいて、ファイルデータを各記憶装置に分散配置する方法であって、各記憶装置が有する記憶容量に比例してファイルデータを配置する比率を決定し、当該決定された比率に従ってファイルデータを各記憶装置に分散して格納することを特徴とする。

【0041】上記のように、第26の発明によれば、各記憶装置が有する記憶容量に比例してファイルデータを配置する比率を決定し、当該決定された比率に従ってフ

ァイルデータを各記憶装置に分散して格納するようにしているので、性能（データの記憶容量）の低い記憶装置がボトルネックとなって性能の高い記憶装置の能力を十分に引き出せなくなるといった事態を防止することができる。

【0042】第27の発明は、複数の記憶装置に対してファイルデータの入出力を行うファイルシステムにおいて、ファイルデータを各記憶装置に分散配置する方法であって、各記憶装置が有する帯域および記憶容量に比例して、ファイルデータを配置する比率を決定し、当該決定された比率に従ってファイルデータを各記憶装置に分散して格納することを特徴とする。

【0043】上記のように、第27の発明によれば、各記憶装置が有する帯域および記憶容量に比例してファイルデータを配置する比率を決定し、当該決定された比率に従ってファイルデータを各記憶装置に分散して格納するようにしているので、性能（データの読出／書込速度および記憶容量）の低い記憶装置がボトルネックとなって性能の高い記憶装置の能力を十分に引き出せなくなるといった事態を防止することができる。

【0044】第28の発明は、複数の記憶装置に対してファイルデータの入出力を行うファイルシステムにおいて、各記憶装置に対するファイルデータの入出力帯域を管理する装置であって、各記憶装置が有する帯域を管理する帯域管理部と、各記憶装置が有する記憶容量を管理する容量管理部と、予め規定されたアルゴリズムに従って、各記憶装置にファイルデータを分配する比率を決定する分配比率決定部と、帯域管理部が保持する管理情報と、分配比率決定部で決定された分配の比率とに基づいて、全記憶装置をまとめた場合の帯域を決定する帯域計算部と、容量管理部が保持する管理情報と、分配比率決定部で決定された分配の比率とに基づいて、全記憶装置をまとめた場合の記憶容量を決定する容量計算部と、帯域計算部で計算された値に基づいて、全記憶装置の入出力帯域の制御を行う全帯域制御部と、容量計算部で計算された値に基づいて、全記憶装置の記憶容量の制御を行う全容量制御部とを備えている。

【0045】上記のように、第28の発明によれば、複数の記憶装置をまとめて入出力帯域および記憶容量を管理するようにしているので、帯域、記憶容量の異なる記憶装置を組み合わせる場合の入出力帯域の管理が容易になる。

【0046】第29の発明は、第28の発明において、分配比率決定部は、各記憶装置が有する帯域に比例して、ファイルデータを配置する比率を決定することを特徴とする。

【0047】第30の発明は、第28の発明において、分配比率決定部は、各記憶装置が有する記憶容量に比例して、ファイルデータを配置する比率を決定することを特徴とする。

【0048】第31の発明は、第28の発明において、分配比率決定部は、各記憶装置が有する帯域および記憶容量に比例して、ファイルデータを配置する比率を決定することを特徴とする。

【0049】上記のように、第29～31の発明によれば、各記憶装置が有する帯域および／または記憶容量に比例してファイルデータを分配する比率を決定するようにしているので、記憶装置全体としての使用効率を最適化することができる。

【0050】第32の発明は、第28の発明において、帯域計算部は、各記憶装置が有する帯域に、分配比率決定部で決定された分配比率の逆数を掛けた値の中で、最も小さい値を、全記憶装置をまとめた場合の帯域とすることを特徴とする。

【0051】第33の発明は、第28の発明において、容量計算部は、各記憶装置が有する記憶容量に、分配比率決定部で決定された分配比率の逆数を掛けた値の中で、最も小さい値を、全記憶装置をまとめた場合の記憶容量とすることを特徴とする。

【0052】第34の発明は、複数の記憶装置に対してファイルデータの入出力を行うファイルシステムであって、ファイルの属性として帯域特性を定義して管理するファイル属性管理部と、ファイルデータの入出力を行う前にファイル属性管理部で管理されているファイルの帯域特性に従って、必要な帯域を獲得する帯域獲得部とを備えている。

【0053】上記のように、第34の発明によれば、ファイルの属性として新たに帯域特性を定義して管理し、この管理された帯域特性に基づいて、ファイルデータの入出力を制御するようにしているので、各ファイルの帯域特性に応じたアクセス制御が行える。

【0054】第35の発明は、第34の発明において、ファイルの属性の帯域特性として、ファイルデータの各記憶装置への分配比率を記述したレイアウトプロファイル情報を用いることを特徴とする。

【0055】第36の発明は、複数の記憶装置に対してファイルデータの入出力を行うファイルシステムであって、ファイルのオープン処理を行うファイルオープン処理部と、ファイルデータの入出力に使用する帯域の獲得を行う帯域獲得部と、ファイルオープン処理部によってオープンされた複数のファイルに対し、帯域獲得部で獲得した帯域を分割して個別的に割り当てることにより、複数のファイルのデータの入出力を同時に行うデータ入出力部とを備えている。

【0056】第37の発明は、複数の記憶装置に対してファイルデータの入出力を行うファイルシステムであって、ファイルのオープン処理を行うファイルオープン処理部と、ファイルデータの入出力に使用する帯域の獲得を行う帯域獲得部と、ファイルオープン処理部によってオープンされた複数のファイルに対し、帯域獲得部で獲

得した1つの帯域を統合的に割り当てることにより、同一の帯域を用いて複数のファイルのデータの入出力を交互に行うデータ入出力部とを備えている。

【0057】上記のように、第36または第37の発明によれば、ファイルのオープンと帯域の獲得とを独立させて任意に組み合わせることにより、データ入出力に使用する帯域の自由度を向上させることができる。

【0058】第38の発明は、サーバから、複数のクライアントに、それぞれ周期的にサービスが提供されるよう、サービスリクエストのスケジューリングを行うスケジューリング装置であって、複数のクライアントからのサービスリクエストをキューイングする第1のキュー手段と、複数のクライアント毎の、実行権を発生させるべき周期を登録する登録手段と、登録手段に登録された周期に関連して、サービスリクエストを実行するための実行権を発生させる実行権発生手段と、第1のキュー手段にキューイングされているサービスリクエストのうち、実行権発生手段が発生させた実行権に対応するクライアントからのサービスリクエストを、サーバに指示して実行させる実行指示手段とを備えている。

【0059】上記のように、第1のキュー手段は、各クライアントからのサービスリクエストをキューイングする。登録手段は、クライアント毎の、実行権を発生させるべき周期を登録する。実行権発生手段は、登録手段に登録された周期に関連して、実行権を発生させる。実行指示手段は、第1のキュー手段にキューイングされているサービスリクエストのうち、実行権発生手段が発生させた実行権に対応するクライアントからのサービスリクエストを、サーバに指示して実行させる。

【0060】このように、各クライアントからのサービスリクエストをキューイングする一方、クライアント毎の、実行権を発生させるべき周期を登録しておいて、その周期に基づいて、実行権を発生させる。そして、キューイングしているサービスリクエストのうち、発生させた実行権に対応するクライアントからのものを、サーバに指示して実行させる。これにより、あるクライアントからのサービスリクエストが送信されてから実行されるまでの時間が、他のクライアントからのサービスリクエストの送信の状況に影響されることがなくなり、従って、送信されたサービスリクエストが実行されるまでの時間が不定となるのを防ぐことができる。

【0061】第39の発明は、第38の発明において、実行権発生手段は、所定の周期で起動され、起動時に、登録手段に登録された、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、実行権を発生させるか否かをクライアント毎に判定し、かつ、実行権を発生させると判定したクライアントの実行権を、一括して発生させることを特徴としている。

【0062】上記のように、第39の発明では、実行権発生手段は、所定の周期で起動され、実行権を発生させ

る動作を行う。この際、実行権発生手段は、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、クライアント毎に、実行権を発生させるか否かを判定し、発生させると判定したクライアントの実行権を、一括して発生させるようにする。これにより、登録手段に登録された、実行権を発生させるべき周期で実行権を発生させるのに比べて、実行権発生手段の処理動作を軽減できる。

【0063】第40の発明は、第39の発明において、実行権発生手段は、登録手段に登録された、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、起動された現時刻から、次の実行権を発生させる時刻までの時間を算定し、算定して得た時間と、自己が起動される周期とを比較することにより、実行権を発生させるか否かを判定することを特徴としている。

【0064】これにより、各クライアントの実行権を発生させるか否かを判定することができる。

【0065】第41の発明は、第40の発明において、実行権発生手段は、起動された現時刻から、次の実行権を発生させる時刻までの時間 $Tofs$ が、自己が起動される周期 $Tsl t$ に対して、 $Tofs < Tsl t$ であるクライアントの実行権を発生させ、さらに、当該クライアントの $Tofs$ に、登録手段に登録された、実行権を発生させるべき周期 $Treq$ を加算し、加算して得た $Tofs$ が、 $Tofs \geq Tsl t$ である場合、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とすることを特徴としている。

【0066】上記のように、第41の発明では、実行権発生手段は、 $Tofs$ が、 $Tsl t$ に対して、 $Tofs < Tsl t$ であるクライアントの実行権を発生させる。さらに、当該クライアントの $Tofs$ に $Treq$ を加算し、加算して得た $Tofs$ が、 $Tofs \geq Tsl t$ である場合、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定し、算定して得た $Tofs$ を、次回起動される際の $Tofs$ とする。これにより、実行権を発生させるべきクライアントを判定して、実行権を発生させることができる。

【0067】第42の発明は、第41の発明において、実行権発生手段は、加算して得た $Tofs$ が、 $Tofs < Tsl t$ である場合には、さらに、実行権を発生させ、かつ当該 $Tofs$ に $Treq$ を加算して $Tsl t$ と比較する動作を繰り返し、加算して得た $Tofs$ が、 $Tofs \geq Tsl t$ となった時点で、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とすることを特徴としている。

【0068】上記のように、第42の発明では、実行権発生手段は、加算して得た $Tofs$ が、 $Tofs < Tsl t$ である場合には、さらに、実行権を発生させると

もに、当該 $Tofs$ に $Treq$ を加算して $Tsl t$ と比較する動作を繰り返す。そして、加算して得た $Tofs$ が、 $Tofs \geq Tsl t$ となった時点で、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とする。これにより、複数の実行権を発生させるべきクライアントをさらに判定して、必要な数の実行権を発生させることができる。

【0069】第43の発明は、第41の発明において、実行権発生手段は、起動された現時刻から、次の実行権を発生させる時刻までの時間 $Tofs$ が、 $Tofs \geq Tsl t$ であるか、または、サービスリクエストが、第1のキュー手段にキューイングされていないクライアントについては、実行権を発生させず、かつ、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とすることを特徴としている。

【0070】上記のように、第43の発明では、実行権発生手段は、 $Tofs$ が、 $Tofs \geq Tsl t$ であるか、または、サービスリクエストが、第1のキュー手段にキューイングされていないクライアントについては、実行権を発生させない。そして、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とする。このように、サービスリクエストがキューイングされていないクライアントについては、実行権を発生させないことにより、サービスリクエストを送信しなかったクライアントにサービスが提供されないようにすることができる。

【0071】第44の発明は、第43の発明において、実行権発生手段は、算定して得た新たな $Tofs$ が、 $Tofs < 0$ である場合には、次回起動される際の $Tofs$ を0とすることを特徴としている。

【0072】これにより、サービスリクエストがキューイングされていないクライアントの $Tofs$ が、繰り返し $Tsl t$ を減算されて小さな値となり、サービスリクエストがキューイングされた時点で当該クライアントの実行権が多数発生されるのを防ぐことができる。

【0073】第45の発明は、第39の発明において、実行指示手段は、所定の周期で起動され、起動された現時刻から、次の実行権が発生される時刻までの時間が、より短いクライアントの実行権を選択することを特徴としている。

【0074】このように、起動された現時刻から、次の実行権が発生される時刻までの時間が、より短いクライアントの実行権を選択することにより、遅延が発生する回数を、より少なくすることができる。

【0075】第46の発明は、第45の発明において、実行権発生手段は、優先度を示す評価指数を、登録手段に登録された、実行権を発生させるべき周期の関数とし

て算定して、発生させた実行権に付与し、実行指示手段は、付与された評価指数に基づいて、実行権発生手段が発生させた実行権のうち、優先されるべき実行権を選択することを特徴としている。

【0076】上記のように、第46の発明では、実行権発生手段は、発生させた実行権に、実行権を発生させるべき周期の関数として算定した、優先度を示す評価指数を付与し、実行指示手段は、付与された評価指数に基づいて、優先されるべき実行権を選択する。これにより、実行権が競合する場合、実行権を発生させるべき周期に関連して、より優先されるべきものを選択することができる。

【0077】第47の発明は、第46の発明において、実行権発生手段は、評価指数 $Nsft$ を、登録手段に登録された実行権を発生させるべき周期 $Treq$ と、自己が起動される周期 $Tslt$ とに対して、関数 $Nsft = f(Treq/Tslt)$ (ただし、 $f(x)$ は x 以上の最小の整数) により算定し、実行指示手段は、付与された評価指数 $Nsft$ がより小さい実行権を選択することを特徴としている。

【0078】これにより、実行権が競合する場合、実行権を発生させるべき周期がより短いものを選択することができる。

【0079】第48の発明は、第47の発明において、実行権発生手段は、登録手段に登録された周期のうち、最も長いものよりは短い周期で、起動されることを特徴としている。

【0080】これにより、全てのクライアントの実行権に、同一の評価指数が付与されることがない。

【0081】第49の発明は、第47の発明において、実行権発生手段は、登録手段に登録された周期のうち、最も短いものよりさらに短い周期で、起動されることを特徴としている。

【0082】これにより、各クライアントの実行権に、より詳細な評価指数が付与される。

【0083】第50の発明は、第46の発明において、実行指示手段は、実行権発生手段が発生させた実行権を、付与された評価指数毎で、かつ発生された順に、キューイングする第2のキュー手段を含み、当該キュー手段にキューイングされている実行権のうち、付与された評価指数がより小さく、かつ、より先に発生されたものを選択することを特徴としている。

【0084】これにより、実行指示手段は、発生された実行権のうち、優先されるべきものを容易に選択することができ、従って、実行指示手段の処理動作を軽減できる。

【0085】第51の発明は、第46の発明において、実行指示手段は、起動時に実行を指示するサービスリクエストの最大数を認識しており、当該最大数に等しい個数だけサービスリクエストの実行を指示した時点で、実

行の指示を停止し、かつ、残された実行権の評価指数を、より優先度の高いものに変更することを特徴としている。

【0086】上記のように、第51の発明では、実行指示手段は、最大数の実行指示を行った時点で、実行指示を停止し、かつ、残された実行権の評価指数を、より優先度の高いものに変更する。これにより、最大数を超えて、実行指示が行われなくようにすることができる。同時に、残された実行権が、長時間、選択されないのを防ぐことができる。

【0087】第52の発明は、第51の発明において、実行指示手段が認識している、起動時に実行を指示するサービスリクエストの最大数は、自己が起動される周期に等しい時間内にサーバが実行可能なサービスリクエストの最大数であることを特徴としている。

【0088】これにより、サーバの処理能力を超えて、実行指示が行われなくようにすることができる。

【0089】第53の発明は、第47の発明において、実行指示手段は、起動時に実行を指示するサービスリクエストの最大数を認識しており、当該最大数に等しい個数だけサービスリクエストの実行を指示した時点で、実行の指示を停止し、かつ、残された実行権の評価指数をディクリメントすることを特徴としている。

【0090】上記のように、第53の発明では、実行指示手段は、最大数の実行指示を行った時点で、実行指示を停止し、かつ、残された実行権の評価指数をディクリメントする。これにより、最大数を超えて、実行指示が行われなくようにすることができる。同時に、残された実行権の評価指数が $Tslt$ 毎にディクリメントされるため、発生されてから、遅くとも $Treq + Tslt$ 後までに、当該実行権が選択されるようにすることができる。

【0091】第54の発明は、第53の発明において、実行指示手段が認識している、起動時に実行を指示するサービスリクエストの最大数は、自己が起動される周期に等しい時間内にサーバが実行可能なサービスリクエストの最大数であることを特徴としている。

【0092】これにより、サーバの処理能力を超えて、実行指示が行われなくようにすることができる。

【0093】第55の発明は、第51の発明において、複数のクライアントからの申告に基づいて、実行権を発生させるべき周期を、登録手段に指示して登録させる登録指示手段をさらに備えることを特徴としている。

【0094】これにより、実行権を発生させるべき周期を、必要に応じて、登録手段に登録させることができる。

【0095】第56の発明は、第55の発明において、登録指示手段は、単位時間当たりに実行指示手段が実行を指示するサービスリクエストの最大数を認識しており、すでに登録されている周期に加えて、さらに他の周

期を登録するよう申告された際、さらに登録されても、発生される実行権の単位時間当たりの平均数が、当該最大数を越えないと判定した場合、登録手段に指示して当該周期を登録させ、さらに登録されれば、越えると判定した場合には、登録させないことを特徴としている。

【0096】上記のように、第56の発明では、登録指示手段は、すでに登録されている周期に加えて、申告された周期がさらに登録されても、実行権発生手段が発生させる実行権の単位時間当たりの平均数が、単位時間当たりに実行指示手段が実行を指示するサービスリクエストの最大数を越えないと判定した場合にのみ、登録手段に指示して、申告された周期を登録させる。これにより、新たな周期を登録した際、発生される実行権の単位時間当たりの平均数が、単位時間当たりに実行指示手段が実行を指示するサービスリクエストの最大数を越えるのを防ぐことができる。

【0097】第57の発明は、複数のサーバから、クライアントに、それぞれ周期的にサービスが提供されるよう、サービスリクエストのスケジューリングを行うスケジューリング装置であって、複数のサーバへのサービスリクエストをキューイングする第1のキュー手段と、複数のサーバ毎の、実行権を発生させるべき周期を登録する登録手段と、登録手段に登録された周期に関連して、サービスリクエストを実行するための実行権を発生させる実行権発生手段と、第1のキュー手段にキューイングされているサービスリクエストのうち、実行権発生手段が発生させた実行権に対応するサーバへのサービスリクエストを、当該サーバに指示して実行させる実行指示手段とを備えている。

【0098】上記のように、第1のキュー手段は、各サーバへのサービスリクエストをキューイングする。登録手段は、サーバ毎の、実行権を発生させるべき周期を登録する。実行権発生手段は、登録手段に登録された周期に関連して、実行権を発生させる。実行指示手段は、第1のキュー手段にキューイングされているサービスリクエストのうち、実行権発生手段が発生させた実行権に対応するサーバのサービスリクエストを、当該サーバに指示して実行させる。

【0099】このように、各サーバへのサービスリクエストをキューイングする一方、サーバ毎の、実行権を発生させるべき周期を登録しておいて、その周期に基づいて、実行権を発生させる。そして、キューイングしているサービスリクエストのうち、発生させた実行権に対応するサーバへのものを、そのサーバに指示して実行させる。これにより、あるサーバへのサービスリクエストが送信されてから実行されるまでの時間が、他のサーバへのサービスリクエストの送信の状況に影響されることがなくなり、従って、送信されたサービスリクエストが実行されるまでの時間が不定となるのを防ぐことができる。

【0100】第58の発明は、第57の発明において、実行権発生手段は、所定の周期で起動され、起動時に、登録手段に登録された、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、実行権を発生させるか否かをサーバ毎に判定し、かつ、実行権を発生させると判定したサーバの実行権を、一括して発生させることを特徴としている。

【0101】上記のように、第58の発明では、実行権発生手段は、所定の周期で起動され、実行権を発生させる動作を行う。この際、実行権発生手段は、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、サーバ毎に、実行権を発生させるか否かを判定し、発生させると判定したサーバの実行権を、一括して発生させるようにする。これにより、登録手段に登録された、実行権を発生させるべき周期で実行権を発生させるのに比べて、実行権発生手段の処理動作を軽減できる。

【0102】第59の発明は、第58の発明において、実行権発生手段は、登録手段に登録された、実行権を発生させるべき周期と、自己が起動される周期とに基づいて、起動された現時刻から、次の実行権を発生させる時刻までの時間を算定し、算定して得た時間と、自己が起動される周期とを比較することにより、実行権を発生させるか否かを判定することを特徴としている。

【0103】これにより、各サーバの実行権を発生させるか否かを判定することができる。

【0104】第60の発明は、第59の発明において、実行権発生手段は、起動された現時刻から、次の実行権を発生させる時刻までの時間 T_{ofs} が、自己が起動される周期 T_{slt} に対して、 $T_{ofs} < T_{slt}$ であるサーバの実行権を発生させ、さらに、当該サーバの T_{ofs} に、登録手段に登録された、実行権を発生させるべき周期 T_{itv} を加算し、加算して得た T_{ofs} が、 $T_{ofs} \geq T_{slt}$ である場合、当該 T_{ofs} に対して、式 $T_{ofs} = T_{ofs} - T_{slt}$ により、新たな T_{ofs} を算定して、次回起動される際の T_{ofs} とすることを特徴としている。

【0105】上記のように、第60の発明では、実行権発生手段は、 T_{ofs} が、 T_{slt} に対して、 $T_{ofs} < T_{slt}$ であるサーバの実行権を発生させる。さらに、当該サーバの T_{ofs} に T_{itv} を加算し、加算して得た T_{ofs} が、 $T_{ofs} \geq T_{slt}$ である場合、当該 T_{ofs} に対して、式 $T_{ofs} = T_{ofs} - T_{slt}$ により、新たな T_{ofs} を算定し、算定して得た T_{ofs} を、次回起動される際の T_{ofs} とする。これにより、実行権を発生させるべきサーバを判定して、実行権を発生させることができる。

【0106】第61の発明は、第60の発明において、実行権発生手段は、加算して得た T_{ofs} が、 $T_{ofs} < T_{slt}$ である場合には、さらに、実行権を発生させ、かつ当該 T_{ofs} に T_{itv} を加算して T_{slt} と

比較する動作を繰り返し、加算して得た $Tofs$ が、 $Tofs \geq Tsl t$ となった時点で、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とすることを特徴としている。

【0107】上記のように、第61の発明では、実行権発生手段は、加算して得た $Tofs$ が、 $Tofs < Tsl t$ である場合には、さらに、実行権を発生させるとともに、当該 $Tofs$ に $Titv$ を加算して $Tsl t$ と比較する動作を繰り返す。そして、加算して得た $Tofs$ が、 $Tofs \geq Tsl t$ となった時点で、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とする。これにより、複数の実行権を発生させるべきサーバをさらに判定して、必要な数の実行権を発生させることができる。

【0108】第62の発明は、第60の発明において、実行権発生手段は、起動された現時刻から、次の実行権を発生させる時刻までの時間 $Tofs$ が、 $Tofs \geq Tsl t$ であるか、または、サービスリクエストが、第1のキュー手段にキューイングされていないサーバについては、実行権を発生させず、かつ、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とすることを特徴としている。

【0109】上記のように、第62の発明では、実行権発生手段は、 $Tofs$ が、 $Tofs \geq Tsl t$ であるか、または、サービスリクエストが、第1のキュー手段にキューイングされていないサーバについては、実行権を発生させない。そして、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl t$ により、新たな $Tofs$ を算定して、次回起動される際の $Tofs$ とする。このように、サービスリクエストがキューイングされていないサーバについては、実行権を発生させないことにより、サービスリクエストを送信されなかったサーバが、サービスを提供するのを防ぐことができる。

【0110】第63の発明は、第62の発明において、実行権発生手段は、算定して得た新たな $Tofs$ が、 $Tofs < 0$ である場合には、次回起動される際の $Tofs$ を0とすることを特徴としている。

【0111】これにより、サービスリクエストがキューイングされていないサーバの $Tofs$ が、繰り返し $Tsl t$ を減算されて小さな値となり、サービスリクエストがキューイングされた時点で当該サーバの実行権が多数発生するのを防ぐことができる。

【0112】第64の発明は、第58の発明において、実行指示手段は、所定の周期で起動され、起動された現時刻から、次の実行権が発生される時刻までの時間が、より短いサーバの実行権を選択することを特徴としている。

【0113】このように、起動された現時刻から、次の実行権が発生される時刻までの時間が、より短いサーバの実行権を選択することにより、遅延が発生する回数を、より少なくすることができる。

【0114】第65の発明は、第64の発明において、実行権発生手段は、優先度を示す評価指数を、登録手段に登録された、実行権を発生させるべき周期の関数として算定して、発生させた実行権に付与し、実行指示手段は、付与された評価指数に基づいて、実行権発生手段が発生させた実行権のうち、優先されるべき実行権を選択することを特徴としている。

【0115】上記のように、第65の発明では、実行権発生手段は、発生させた実行権に、実行権を発生させるべき周期の関数として算定した、優先度を示す評価指数を付与し、実行指示手段は、付与された評価指数に基づいて、優先されるべき実行権を選択する。これにより、実行権が競合する場合、実行権を発生させるべき周期に関連して、より優先されるべきものを選択することができる。

【0116】第66の発明は、第65の発明において、実行権発生手段は、評価指数 $Nsft$ を、登録手段に登録された実行権を発生させるべき周期 $Titv$ と、自己が起動される周期 $Tsl t$ とに対して、関数 $Nsft = f(Titv / Tsl t)$ （ただし、 $f(x)$ は x 以上の最小の整数）により算定し、実行指示手段は、付与された評価指数 $Nsft$ がより小さい実行権を選択することを特徴としている。

【0117】これにより、実行権が競合する場合、実行権を発生させるべき周期がより短いものを選択することができる。

【0118】第67の発明は、第66の発明において、実行権発生手段は、登録手段に登録された周期のうち、最も長いものよりは短い周期で、起動されることを特徴としている。

【0119】これにより、全てのサーバの実行権に、同一の評価指数が付与されることがない。

【0120】第68の発明は、第66の発明において、実行権発生手段は、登録手段に登録された周期のうち、最も短いものよりさらに短い周期で、起動されることを特徴としている。

【0121】これにより、各サーバの実行権に、より詳細な評価指数が付与される。

【0122】第69の発明は、第65の発明において、実行指示手段は、実行権発生手段が発生させた実行権を、付与された評価指数毎で、かつ発生された順に、キューイングする第2のキュー手段を含み、当該キュー手段にキューイングされている実行権のうち、付与された評価指数がより小さく、かつ、より先に発生されたものを選択することを特徴としている。

【0123】これにより、実行指示手段は、発生された

実行権のうち、優先されるべきものを容易に選択することができ、従って、実行指示手段の処理動作を軽減できる。

【0124】第70の発明は、第65の発明において、実行指示手段は、起動時に実行を指示するサービスリクエストの最大数を認識しており、当該最大数に等しい個数だけサービスリクエストの実行を指示した時点で、実行の指示を停止し、かつ、残された実行権の評価指数を、より優先度の高いものに変更することを特徴としている。

【0125】上記のように、第70の発明では、実行指示手段は、最大数の実行指示を行った時点で、実行指示を停止し、かつ、残された実行権の評価指数を、より優先度の高いものに変更する。これにより、最大数を超えて、実行指示が行われないようにすることができるとともに、残された実行権が、長時間、選択されないのを防ぐことができる。

【0126】第71の発明は、第70の発明において、実行指示手段が認識している、起動時に実行を指示するサービスリクエストの最大数は、自己が起動される周期に等しい時間内に、全サーバからクライアントに送信可能な最大の情報量に対応するサービスリクエストの個数であることを特徴としている。

【0127】これにより、全サーバからクライアントに送信可能な最大の情報量を超えるような数の実行指示が行われないようにすることができる。

【0128】第72の発明は、第66の発明において、実行指示手段は、起動時に実行を指示するサービスリクエストの最大数を認識しており、当該最大数に等しい個数だけサービスリクエストの実行を指示した時点で、実行の指示を停止し、かつ、残された実行権の評価指数をディクリメントすることを特徴としている。

【0129】上記のように、第72の発明では、実行指示手段は、最大数の実行指示を行った時点で、実行指示を停止し、かつ、残された実行権の評価指数をディクリメントする。これにより、最大数を超えて、実行指示が行われないようにすることができるとともに、残された実行権の評価指数がT s l t毎にディクリメントされるため、発生してから、遅くともT r e q+T s l t後までに、当該実行権が選択されるようにすることができる。

【0130】第73の発明は、第72の発明において、実行指示手段が認識している、起動時に実行を指示するサービスリクエストの最大数は、自己が起動される周期に等しい時間内に、全サーバからクライアントに送信可能な最大の情報量に対応するサービスリクエストの個数であることを特徴としている。

【0131】これにより、全サーバからクライアントに送信可能な最大の情報量を超えるような数の実行指示が行われないようにすることができる。

【0132】第74の発明は、第70の発明において、クライアントからの申告に基づいて、実行権を発生させるべき周期を、登録手段に指示して登録／更新させる登録指示手段をさらに備えることを特徴としている。

【0133】これにより、実行権を発生させるべき周期を、必要に応じて、登録手段に登録／更新させることができる。

【0134】第75の発明は、第74の発明において、登録指示手段は、単位時間あたりに実行指示手段が全サーバに実行を指示するサービスリクエストの最大数を認識しており、すでに登録されている周期に加えて、さらに他のサーバの周期を登録するよう申告された際、さらに登録されても、発生される、全サーバの実行権の単位時間当たりの平均数が、当該最大数を超えないと判定した場合、登録手段に指示して当該周期を登録させ、さらに登録されれば、越えると判定した場合には、登録させないことを特徴としている。

【0135】上記のように、第75の発明では、登録指示手段は、すでに登録されている周期に加えて、申告された周期がさらに登録されても、実行権発生手段が発生させる、全サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手段が全サーバに実行を指示するサービスリクエストの最大数を超えないと判定した場合にのみ、登録手段に指示して、申告された周期を登録させる。これにより、申告されたサーバの周期を登録した際、発生される、全サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手段が全サーバに実行を指示するサービスリクエストの最大数を超えるのを防ぐことができる。

【0136】第76の発明は、第74の発明において、登録指示手段は、単位時間あたりに実行指示手段が各サーバに実行を指示するサービスリクエストの最大数を認識しており、すでに登録されている周期に加えて、さらに他のサーバの周期を登録するよう申告された際、さらに登録されても、発生される、当該サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手段が当該サーバに実行を指示するサービスリクエストの最大数を超えないと判定した場合、登録手段に指示して当該周期を登録させ、さらに登録されれば、越えると判定した場合には、登録させないことを特徴としている。

【0137】上記のように、第76の発明では、登録指示手段は、すでに登録されている周期に加えて、申告されたサーバの周期がさらに登録されても、実行権発生手段が発生させる、当該サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手段が当該サーバに実行を指示するサービスリクエストの最大数を超えないと判定した場合にのみ、登録手段に指示して、申告された周期を登録させる。これにより、申告されたサーバの周期を登録した際、発生される、当該サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行

指示手段が当該サーバに実行を指示するサービスリクエストの最大数を越えるのを防ぐことができる。

【0138】第77の発明は、第76の発明において、登録指示手段が認識している、単位時間あたりに実行指示手段が各サーバに実行を指示するサービスリクエストの最大数は、単位時間あたりに各サーバが実行可能なサービスリクエストの最大数であることを特徴としている。

【0139】これにより、申告されたサーバの周期を登録した際、発生される、当該サーバの実行権の単位時間当たりの平均数が、単位時間あたりに当該サーバが実行可能なサービスリクエストの最大数を越えるのを防ぐことができる。

【0140】第78の発明は、第74の発明において、登録指示手段は、単位時間あたりに実行指示手段が全サーバに実行を指示するサービスリクエストの最大数と、単位時間あたりに実行指示手段が各サーバに実行を指示するサービスリクエストの最大数とを認識しており、すでに登録されている、あるサーバの実行権を発生させるべき周期 T_{itv} を、さらに周期 T_{req} を加えて実行権を発生させるものに更新するよう、申告を受けた際、更新されても、発生される、全サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手段が全サーバに実行を指示するサービスリクエストの最大数を越えないと判定し、かつ、発生される、当該サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手段が当該サーバに実行を指示するサービスリクエストの最大数を越えないと判定した場合、登録手段に指示して、すでに登録されている T_{itv} を、式 $T_{itv} = 1 / (1 / T_{itv} + 1 / T_{req})$ により算定される新たな T_{itv} に更新させ、更新されれば、越えるのと判定した場合には、更新させないことを特徴としている。

【0141】上記のように、第78の発明では、あるサーバの T_{itv} を、さらに周期 T_{req} を加えて実行権を発生させるものに更新するよう、申告を受けた際、更新されても、発生される、全サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手段が全サーバに実行を指示するサービスリクエストの最大数を越えず、かつ、発生される、当該サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手段が当該サーバに実行を指示するサービスリクエストの最大数を越えないと判定した場合にのみ、登録手段に指示して、当該サーバの T_{itv} を更新させる。これにより、当該サーバの T_{itv} を更新した際、発生される、全サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手段が全サーバに実行を指示するサービスリクエストの最大数を越えないようにすることができるとともに、発生される、当該サーバの実行権の単位時間当たりの平均数が、単位時間あたりに実行指示手

段が当該サーバに実行を指示するサービスリクエストの最大数を越えないようにすることができる。

【0142】第79の発明は、第74の発明において、登録指示手段は、すでに登録されている、あるサーバの実行権を発生させるべき周期 T_{itv} を、周期 T_{req} を除いて実行権を発生させるものに更新するよう申告を受けた際、登録手段に指示して、すでに登録されている T_{itv} を、式 $T_{itv} = 1 / (1 / T_{itv} - 1 / T_{req})$ により算定される新たな T_{itv} に更新させることを特徴としている。

【0143】これにより、すでに登録されている、実行権を発生させるべき周期を、周期 T_{req} を除いて発生させるものに更新するよう申告を受けた際、すでに登録されている周期を、新たなものに更新することができる。

【0144】第80の発明は、有線または無線の伝送路を介して複数の端末と接続されており、いずれかの端末からデータの読み出し要求が送られてきたとき、内部に蓄積されたデータの中から対応するデータを検索して読み出し、読み出したデータを読み出し要求を発した端末へと送信するサーバ装置であって、種々のデータを記憶しており、端末からの読み出し要求に応じて、ランダムアクセスが可能なデータ記憶部と、データ記憶部と各端末との間でデータの入出力を制御する入出力部と、入出力部とデータ記憶部との間でデータを交換する交換部と、端末から要求されたデータを、データ記憶部から読み出して当該端末へと出力する際に必要となるリソースを管理するリソース管理部とを備え、リソース管理部は、サーバ装置全体のリソースを、連続性を保証する必要があるデータ（優先データ）を端末へと出力するために用いられるリソース（優先リソース）と、連続性を保証する必要のないデータ（非優先データ）を端末へと出力するために用いられるリソース（非優先リソース）とに分離して管理することを特徴とする。

【0145】上記のように、第80の発明によれば、システム全体のリソースを、優先リソースと非優先リソースとに分離して管理するようにしているので、非優先データの読み出しに使用するリソースを、優先リソースとは独立して確保することができる。その結果、非優先データの読み出し量が増加しても、優先データの出力に影響を与えることがない。

【0146】第81の発明は、第80の発明において、各端末は、複数のグループに分割されており、入出力部は、端末のグループ毎に設けられた複数のI/Oユニットを含み、各I/Oユニットは、対応するグループの端末から送られてくる優先データの読み出しを要求するコマンドに対する入出力制御を実行する優先入出力制御部と、対応するグループの端末から送られてくる非優先データの読み出しを要求するコマンドに対する入出力制御を実行する非優先入出力制御部とを含み、データ記憶部

は、複数の記憶装置と、各記憶装置と接続され、当該記憶装置からのデータの読み出しを制御する複数のMSFSとを含み、非優先入出力制御部は、対応するグループの端末から送られてくる非優先データの読み出しを要求するコマンドを一時的に格納するコマンドキューと、コマンドキューからコマンドを取り出すコマンドキュー制御部と、コマンドキュー制御部が取り出したコマンドを、対応する宛先のMSFSに送出するコマンド送出部と、MSFSによって記憶装置から読み出された非優先データを一時的に格納するデータバッファ部と、データバッファ部に格納された非優先データを端末に送出するデータ送出部とを含んでいる。

【0147】第82の発明は、第81の発明において、コマンドキューは、対応するグループに属する複数の端末のそれぞれに対応して設けられており、それぞれが対応する端末からのコマンドを到着順に格納し、コマンドキュー制御部は、複数のコマンドキューの中からコマンドを格納している1つのコマンドキューを選択すると共に、その先頭に格納されているコマンドおよびそれと同じMSFSに送出されるコマンドを取り出してコマンド送出部に出力するコマンド取り出し処理を実行し、コマンド送出部は、コマンドキュー制御部から受け取ったコマンドを、それを実行すべきMSFSに送出し、送出したコマンドに対する全ての読み出しデータを、データバッファ部がMSFSから受信したことを確認すると、受信完了通知をコマンドキュー制御部に送出し、コマンドキュー制御部は、コマンド送出部から受信完了通知を受け取る毎に、コマンド取り出し処理を繰り返し実行することを特徴とする。

【0148】上記のように、第82の発明では、コマンドキュー制御部は、コマンド送出部にコマンドを出力後、出力したコマンドに対応する全てのデータを、データバッファ部が受信するまでは、次のコマンドをコマンド送出部に出力しないことになる。これによって、I/Oユニットは、非優先データの受信に関して、同時に複数のMSFSと通信することがなくなり、非優先データ同士のセルの衝突を考慮する必要がない。一方、MSFSは、I/Oユニットが持つ非優先リソースの通信帯域の全てを使用して高速にデータを送信することが可能となる。

【0149】第83の発明は、第82の発明において、コマンドキュー制御部は、1回のコマンド取り出し処理において、1つのコマンドキューから取り出すコマンドの数を、予め定めた一定個数以下に制限することを特徴とする。

【0150】1つのMSFSに対してほぼ同時に多数のコマンドが送出されると、コマンドを受信したMSFSで輻輳が生じるため、上記第82の発明では、1つのコマンドキューから取り出されてコマンド送出部に送出されるコマンド数を、予め定めた一定個数以下に制限する

ようにしている。

【0151】第84の発明は、第82の発明において、非優先入出力制御部は、データバッファ部の使用状態を管理するバッファ管理部をさらに含み、コマンドキュー制御部は、バッファ管理部が管理するデータバッファの使用状態に基づき、コマンド取りだし処理のために選択するコマンドキューを決定することを特徴とする。

【0152】第85の発明は、第84の発明において、バッファ管理部には、各端末毎にデータバッファ部に格納されている未送信データ数をカウントする複数の使用量カウンタが設けられており、コマンドキュー制御部は、コマンド取り出し処理を実行する前にバッファ管理部の使用量カウンタを参照し、未送信データ数が最も小さい端末に対応するコマンドキューを、コマンド取り出し処理の対象として選択することを特徴とする。

【0153】上記のように、第85の発明では、バッファ管理部が管理するデータバッファ部の使用状況に基づいて、コマンドが蓄えられているコマンドキューの中から、データバッファ部に格納されているデータ量が最も少ない端末に対応するコマンドキューを選択し、当該選択したコマンドキューに蓄えられているコマンドの送出処理を優先的に実行するようにしているので、複数の端末が共通に使用している非優先リソースを、1つの端末が占有して使用することが少なくなり、サーバ装置が各端末に対して有する通信能力を最大限に活用することができる。

【0154】第86の発明は、第82の発明において、コマンド送出部は、MSFSによって読み出される非優先データを格納するための領域をデータバッファ部に予約した後に、コマンドキュー制御部から受け取ったコマンドをMSFSに送出し、データバッファ部に非優先データを格納するための領域が予約できない場合には、予約できるまでMSFSに対するコマンドの送出を停止することを特徴とする。

【0155】上記のように、第86の発明では、コマンド送出部は、データバッファ部内にデータを格納するための領域を予約した後に、非優先データの読み出しを要求するコマンドをMSFSに送出するようにしている。すなわち、コマンド送出部は、格納領域を予約できない場合には、予約できるまで、非優先データの読み出しを要求するコマンドをMSFSに送出しない。これによって、MSFSからのデータ受信時に、データバッファ部内に格納領域がなくてデータが廃棄されることを防止することができる。

【0156】第87の発明は、第81の発明において、コマンドキュー制御部は、I/Oユニットが端末にデータを出力するための非優先リソースとしての通信帯域と、I/Oユニットがデータ記憶部から読み出された非優先データを受信するための非優先リソースとしての通信帯域との内、小さい方の帯域に基づいて、コマンド送

出部に対してコマンドを出力する周期を決定することを特徴とする。

【0157】上記のように、第87の発明では、コマンドキュー制御部は、I/Oユニットが端末にデータを出力するための非優先リソースとしての通信帯域と、I/Oユニットがデータ記憶部から読み出された非優先データを受信するための非優先リソースとしての通信帯域との内、小さい方の帯域に基づいて、コマンド送出部に対してコマンドを出力する周期を決定するようにしている。これによって、MSFSから送られてくるデータによって使用される平均帯域が、交換部がMSFSと通信するための非優先リソースとしての通信帯域と等しくなるか、またはそれよりも小さくなるため、MSFSのデータ送信による交換部の輻輳を抑制する効果が期待できる。さらに、MSFSから送られてくるデータによって使用される平均帯域が、I/Oユニットが端末にデータを出力するための非優先リソースとしての通信帯域とも等しくなるか、またはそれよりも小さくなるため、I/Oユニット内のデータバッファ部に蓄積されるデータ量が增大していくことを抑えることができる。

【0158】第88の発明は、第81の発明において、コマンド送出部は、各MSFSに対応して設けられ、コマンドキュー制御部から受け取ったコマンドを、MSFS別に格納する複数の送出キューと、各送出キューのそれぞれに対応して設けられ、対応する送出キューからMSFSへコマンドを送出する毎にインクリメントされ、対応するMSFSからの読み出しデータがデータバッファ部に到着する毎にデクリメントされることにより、データバッファ部への非優先データの未到着数をカウントする複数の処理数カウンタとを含み、コマンド送出部は、処理数カウンタによってカウントされる未到着数が所定のしきい値を超えないように、各送出キューからMSFSに対して送出されるコマンドの数を制限することを特徴とする。

【0159】上記のように、第88の発明では、コマンド送出部は、注目している送出キューに対応する処理数カウンタのカウント値が、予め定められたしきい値未満である場合にのみ、当該注目している送出キュー内のコマンドをMSFSに送信するようにしているので、1つのMSFSに対して読み出し処理が集中し、当該MSFS内でコマンドが滞留するのを避けることができる。

【0160】第89の発明は、第80の発明において、リソース管理部は、リソースとして、交換部の通信帯域と、データ記憶部がデータを読み出す帯域と、入出力部が各端末にデータを出力するための通信帯域とを管理していることを特徴とする。

【0161】第90の発明は、第80の発明において、交換部は、ATMスイッチで構成されており、ATMスイッチの非優先リソースとしての通信帯域が、ABRモードで確保されることを特徴とする。

【0162】上記のように、第90の発明では、非優先リソースとしてのATMスイッチの通信帯域をABRモードを用いて確保し、データを読み出したMSFSは、ABRモードで確保している通信帯域を用いてデータを送信するようにしているので、ATMスイッチにおいて輻輳が生じた場合でも、即座にMSFSからの送信レートが下がり、輻輳が解消される。従って、I/Oユニットは、常に正常にデータを受信できる。

【0163】第91の発明は、第80の発明において、リソース管理部は、予め固定的に、サーバ装置全体のリソースの内、一部を優先リソースとして割り当て、残りの部分を非優先リソースとして割り当てることを特徴とする。

【0164】第92の発明は、第80の発明において、リソース管理部は、端末からの優先データの読み出し要求に応じて、サーバ装置全体のリソースの内、優先リソースとして割り当てる量を変化させることを特徴とする。

【0165】上記のように、第92の発明では、端末からの優先データの読み出し要求に応じて、サーバ装置全体のリソースの内、優先リソースとして割り当てる量を変化させるようにしているので、優先データの読み出し要求の増減に応じて、リソースの割り当てを柔軟に変えて、有限なリソースを効率的に活用することができる。

【0166】第93の発明は、第92の発明において、リソース管理部は、サーバ装置全体のリソースの内、優先リソースとして割り当てられていないリソースを全て非優先リソースとして割り当て、優先リソースを増やす必要が生じた場合、非優先リソースが所定のリソース量以下にならない範囲で、当該非優先リソースとして割り当てられているリソースから優先リソースを割り当てることを特徴とする。

【0167】上記のように、第93の発明では、優先リソースとして使用されていないリソースを全て非優先リソースとして管理し、優先リソースを新たに増やす必要が生じた場合には、非優先リソースが所定のリソース量（最低帯域量）以下にならない条件下で、非優先リソースから優先リソースを割り当てるようにしているので、非優先データの読み出しが増加しても、優先データの入出力が影響を受けることがなく、しかも、使用されていないリソースが全て非優先リソースとして使用されるため、非優先データの読み出しが効率的に実行される。

【0168】第94の発明は、利用者側からのデータ送出要求に回答してデータ取得要求をデータ記憶装置に発行し、当該データ取得要求に従ってデータ記憶装置から取得したデータを利用者側に送出するファイルシステムであって、データ送出要求によって指定された位置から順番に、かつ指定された速度で連続的にデータ記憶装置からデータを取得するデータ取得部と、データ取得部が

データ記憶装置から取得したデータを格納するバッファと、バッファに格納されたデータの操作とデータ量の管理とを行うバッファ管理部と、バッファに格納されているデータを、データ送出要求によって指定された順番に、かつ指定された伝送速度で連続的に読み出して利用者側に送出するデータ送出部と、利用者側からのポーズ要求およびポーズ解除要求を受けて、データ送出部、データ取得部およびバッファ管理部を制御するポーズ制御部とを備え、ポーズ制御部は、利用者側からのポーズ要求を受けた場合、直ちにデータ送出を停止するようデータ送出部を制御すると共に、直ちにデータ取得を停止するようデータ取得部を制御し、利用者側からのポーズ解除要求を受けた場合、ポーズ処理以降にバッファに格納されているデータの先頭から順番にデータ送出を再開するようデータ送出部を制御すると共に、ポーズ処理以降にバッファに格納されているデータの最後の次のデータから順番にデータ取得を再開するようデータ取得部を制御することを特徴とする。

【0169】上記のように、第94の発明によれば、利用者側からポーズ解除要求を受けると、最初の内は、ポーズ処理以降にバッファに格納されているデータを利用者側に送出するようにしているため、ポーズ解除時にデータ送出が再開されるまでの時間を、新たにデータ送出要求を行う場合に比べて短縮化することができる。

【0170】第95の発明は、第94の発明において、ポーズ制御部は、利用者側からのポーズ解除要求を受けた場合、データ取得部にデータ取得を再開させるデータ取得再開時刻 t_{rd} と、ポーズ処理以降にバッファに格納されているデータ量 N と、ポーズ解除後のデータ伝送速度 V と、データ取得部がデータ取得に必要とする時間 D とを、次式 (a) に代入することにより、データ送出部にデータ送出を再開させるデータ送出再開時刻 t_{sd} を算出することを特徴とする。

$$t_{sd} = t_{rd} + D - N/V \quad \dots (a)$$

【0171】上記のように、第95の発明によれば、ポーズ解除時にデータ送出を再開する時刻を、式 (a) によって算出するようにしているため、少なくともポーズ解除後にデータ取得部が取得する先頭のデータがバッファに格納されるまでの間は、ポーズ処理以降にバッファに格納されているデータが利用者に送出される。その結果、ポーズ解除時のデータ送出再開までの時間を、新たにデータ送出要求を行う場合に比べて短縮化することができる。また、ポーズ前後のデータ伝送速度が変化した場合においても、送出データがバッファ中になくなって送出データが途切れることを防止でき、データの連続送出を実現することができる。

【0172】第96の発明は、第94の発明において、ポーズ制御部は、利用者側からのポーズ解除要求を受けた場合、データ送出部にデータ送出を再開させるデータ送出再開時刻 t_{sd} と、ポーズ処理以降にバッファに格納

されているデータ量 N と、ポーズ解除後のデータ伝送速度 V と、データ取得部がデータ取得に必要とする時間 D とを、次式 (b) に代入することにより、データ取得部にデータ取得を再開させるデータ取得再開時刻 t_{rd} を算出することを特徴とする。

$$t_{rd} = t_{sd} + N/V - D \quad \dots (b)$$

【0173】上記のように、第96の発明によれば、ポーズ解除時にデータ取得を再開する時刻を、式 (b) によって算出するようにしているため、少なくともポーズ解除後にデータ取得部が取得する先頭のデータがバッファに格納されるまでの間は、ポーズ処理以降にバッファに格納されているデータが利用者に送出される。その結果、ポーズ解除時のデータ送出再開までの時間を、新たにデータ送出要求を行う場合に比べて短縮化することができる。また、ポーズ前後のデータ伝送速度が変化した場合においても、送出データがバッファ中になくなって送出データが途切れることを防止でき、データの連続送出を実現することができる。

【0174】第97の発明は、第94の発明において、ポーズ制御部は、利用者側からのポーズ解除要求を受けた場合、ポーズ処理以降にバッファに格納されているデータ量が、データ取得部がデータ取得に必要とする時間中にポーズ解除後のデータ伝送速度で送出可能な限界データ量よりも大きいかなんかを判断し、判断の結果、データ量が限界データ量よりも大きい場合は、当該限界データ量を越える部分のデータをバッファから削除するようにバッファ管理部を制御し、それによってポーズ解除後にデータ取得部が取得するデータの先頭位置を決定することを特徴とする。

【0175】上記のように、第97の発明によれば、ポーズ解除時に限界データ量を越える部分のデータをバッファから削除するようにしているため、バッファは限界データ量以上のバッファ容量を持つ必要がない。

【発明の実施の形態】 (第1の実施形態) 図1は、本発明の第1の実施形態に係るファイルシステムの構成を示すブロック図である。図1において、本ファイルシステムは、交換装置1と、ファイルマネージャ2と、複数のサーバユニット3とを備えている。

【0176】ファイルマネージャ2は、ファイルシステム全体を管理する。各サーバユニット3は、ユーザからのファイルの読み出しまたは書き込みの要求を受け付け、ファイルシステムが管理する記憶装置から読み出したデータやメッセージをユーザに対して送信する。各サーバユニット3間、および各サーバユニット3とファイルマネージャ2との間では、交換装置1を介してデータがやりとりされる。当該ファイルシステムが管理するデータは、各サーバユニット3内に存在する記憶装置に格納されている。

【0177】図2は、各サーバユニット3におけるファイルアクセスに関連する部分の構成を示すブロック図で

ある。図2において、各サーバユニット3は、外部メッセージインターフェース部101と、リソース割り当て部102と、スロット割り当て部103と、アドレス検索部104と、要求送出部105と、外部ネットワークI/O制御部106と、内部メッセージインターフェース部107と、記憶装置108と、記憶制御部109と、内部ネットワークI/O制御部110と、申告値管理部111とを含む。

【0178】外部メッセージインターフェース部101は、外部とやりとりされるメッセージを識別し、サーバユニット内の各機能ブロックに対して処理の指示を与える。リソース割り当て部102は、ファイルシステム内のリソースの割り当てを行う。スロット割り当て部103は、サーバユニット3で管理しているスロットの割り当てを行う。なお、スロットについての説明は後述する。アドレス検索部104は、要求されたデータの格納位置を検索する。要求送出部105は、固定長のデータブロックの読み出しを要求するブロック読み出し要求や、書き込みを要求するブロック書き込み要求の送出を指示する。内部メッセージインターフェース部107は、ファイルシステム内部のファイルマネージャ2や、他のサーバユニットに対するメッセージの送受信を制御する。記憶装置108は、データブロックや、ファイル毎のデータの格納位置情報を記憶する。記憶制御部109は、記憶装置108へのアクセスを制御する。内部ネットワークI/O制御部110は、ブロック読み出し要求により要求されたデータブロックを交換装置1に送出するタイミングを制御する。申告値管理部111は、外部から到着したメッセージのパラメータが妥当なものかどうかをチェックする。

【0179】図3は、ファイルマネージャ2におけるファイルアクセスに関連する部分の構成を示すブロック図である。図3において、ファイルマネージャ2は、メッセージインターフェース部201と、リソース管理部202とを含む。メッセージインターフェース部201は、交換装置1を介して各サーバユニット3とやりとりされるメッセージを識別し、ファイルマネージャ2内の機能に対して処理の指示を与える。リソース管理部202は、ファイルシステムにおける、記憶装置の読み出し帯域、交換装置の入出力帯域の容量と使用量とを管理する。

【0180】ここで、ファイルシステムには、外部からファイルアクセスに関連して、以下のメッセージが到着する。

- (1) GetResource
- (2) FileOpen
- (3) ReadBlock
- (4) WriteBlock

【0181】上記メッセージを受け取ったときの、ファイルマネージャ2および各サーバユニット3での、ファ

イルアクセスに関連する部分の動作について、以下に説明する。

【0182】GetResourceメッセージが到着すると、サーバユニット3内の外部メッセージインターフェース部101は、リソース割り当て部102にその旨を通知する。リソース割り当て部102は、ファイルマネージャ2に対するリソース割り当て要求の送出を、内部メッセージインターフェース部107に指示する。リソース割り当て要求がファイルマネージャ2に到着すると、リソース管理部202では、リソース割り当て要求の受け付け可否を判断し、その結果をサーバユニット3に送出する。サーバユニット3内のリソース割り当て部102では、割り当て可能の判断結果を受け取ると、リソースの予約を行う。

【0183】FileOpenメッセージが到着すると、サーバユニット3は、ファイル内のデータの格納位置情報を記憶装置108から取得し、ファイルアクセス可能な状態に移行する。

【0184】ReadBlockメッセージが到着すると、サーバユニット3内のスロット割り当て部103は、予約したリソースを割り当てて、要求されたデータの読み出しを行う。

【0185】WriteBlockメッセージが到着すると、サーバユニット3内のスロット割り当て部103は、予約したリソースを割り当てて、要求されたデータの書き込みを行う。

【0186】図4は、リソースの予約および割り当てと、データ読み出し動作との関係を示した図である。以下には、この図4を参照しつつ、リソース予約および割り当てを行ってデータの読み出し、書き込みを行うファイルシステムの動作について説明する。まず始めに、リソースの予約および割り当てに用いるMS到着スロットおよび書き込み用スロットについて説明した後、ファイルシステムの動作について説明する。

【0187】(1) MS到着スロット

本ファイルシステムでは、数10msecおきに区切られた「MS到着スロット」という概念を用いる。図4において、線801は、時間をMS到着スロットに区切って管理していることを示している。また、データの読み出しに際しては、要求されたデータを固定長のデータブロック(MSブロック)に区切り、MS毎にブロック読み出し要求を記憶装置108または他のサーバユニットに送出し、読み出しを行う。あるサーバユニットが他のサーバユニットに対してブロック読み出し要求を送出する場合には、MS到着スロットを割り付ける。1つのスロットに割り付けることができるブロック読み出し要求の個数には、最大値が設定されている。本ファイルシステムでは、ユーザへのストリーム提供を保証するために、ブロック読み出し要求を受け付けたユーザについて、予めMS到着スロットを予約する。ブロック読み出

し要求に割り付けるMS到着スロットは、予約されたMS到着スロットの中から選定する。MS到着スロットを予約する動作は、一定時間おきにまとめて行う。「MS到着スロット」を形成するパラメータを以下に示す。

T_{slot} : スロットの時間間隔

N_{max1} : スロットに割り当て可能な要求数

T_{rsvMS} : 到着スロット予約動作の間隔

【0188】図4において、参照番号802は、 T_{slot} の間隔を示している。図5は、MS到着スロットの状態を管理するMS到着スロット管理テーブルの一例を示している。図5に示したように、MS到着スロット管理テーブルは、スロットを識別する「スロット番号」、スロットの予約の有無および予約しているユーザを識別する「予約ユーザ番号」、ブロック読み出し要求を割り付けているかどうかおよび割り当てている要求を識別する「割り当て要求番号」、非優先のブロック読み出し要求を割り付けてよいかどうかを識別する「非優先フラグ」といった項目を情報として持つ。GetResourceメッセージによってリソースが割り当てられたユーザは、予約したスロットを利用する。また、リソース割り当てなしにブロックの読み出しを要求したユーザに対しては、ブロックの読み出しをする場合には、非優先フラグを参照して、割り当てることが可能なスロットを割り当てる。

【0189】(2) 書き込み用スロット

ブロックの書き込みに際しては、「書き込み用スロット」という概念が用いられる。各サーバユニット3において、書き込み用スロットは、宛先サーバユニット毎に設けられている。外部からWriteBlockコマンドと共に到着するデータは、固定長のデータブロック(MSブロック)に分割され、記憶装置108の空き領域リストに従って宛先が決定され、ブロック書き込み要求として送出される。ここで、ブロック書き込み要求には、宛先となるサーバユニットの書き込み用スロットを必ず割り付ける。書き込み用スロットの状態を管理する書き込み用スロット管理テーブルの一例を図6に示す。図6に示したように、書き込み用スロット管理テーブルは、宛先サーバユニットを識別する「宛先ユニット番号」、書き込み用スロットを特定する「スロット番号」、スロットに割り当てた要求の有無と割り当てた要求を識別する「割り当て要求識別番号」といった項目を情報として持つ。

【0190】(3) ユーザ毎のリソース管理

サーバユニット3では、GetResourceを受け付けているユーザを、Read/Write別に、確保

$$D_{lim} = D_{net} + D_{disk_min} + 2 \times T_{max} \dots (1)$$

上式(1)において、交換装置1での遅延許容値を D_{net} 、記憶装置108での遅延限界値の最小値を D_{disk_min} とする。また、 T_{max} と T を比較した場合に、小さい方を T_{min} とする。

した帯域が大きいものより小さいものの順に「ストリーム管理テーブル」で管理している。確保した帯域により、ユーザは、単数または複数のファイルを読み出すことができる。図7は、ストリーム管理テーブルの一例を示している。このストリーム管理テーブルは、各サーバユニット3内のリソース割り当て部102で管理されている。「ストリーム管理テーブル」は、ユーザ毎のリソースの獲得状況を管理するために、図7に示したような項目を情報として持つ。図7において、「リソース記述子」は、GetResourceメッセージ受け付け毎に割り付けられる識別子である。「ストリーム管理テーブル」は、この「リソース記述子」毎に、読み出し用か書き込み用かを識別する「読み出し・書き込みの別」と、要求帯域を提供するために必要となるスロットの平均予約間隔とを記憶している。さらに、「ストリーム管理テーブル」は、確保した帯域を記憶する「確保されたファイルシステム内の内部帯域」、確保した帯域のうち現在使っていない帯域量を記憶する「現在使用可能なファイルシステム内部帯域」といった項目を情報として持つ。

【0191】(4) リソース獲得受け付け制御

リソース獲得受け付け制御は、外部からのGetResourceメッセージの受け付け可否を判断し、判断結果を外部に返す制御である。受け付け可否の判断は、メッセージによって申告されるパラメータに基づいて、リソース管理部202にて行われる。GetResourceメッセージは、以下の申告パラメータを持つ。

Read/Writeの別

ブロックの最小Read/Write間隔 T

ReadBlock到着から、ブロック送出までの遅延要求値 D_{lim}

【0192】図8は、図3に示すリソース管理部202のより詳細な構成を示すブロック図である。図8において、リソース管理部202は、申告パラメータ処理部401と、書き込み帯域算出部402と、受付判断部403と、リソース記憶部404とを含む。申告パラメータ処理部401は、コマンドによって申告されるパラメータを、受け付け判断に必要な数値に置き換える。書き込み帯域算出部402は、書き込み用の帯域を算出する。受付判断部403は、メッセージの受付の可否を判断する。リソース記憶部404は、現在のリソースの使用状況と、提供可能なリソースの最大値とを記憶する。

【0193】申告パラメータ処理部401では、まず、遅延要求値 D_{lim} に基づき、次式(1)を満たす値 T_{max} を求める。

【0194】申告パラメータ処理部401は、 T_{min} の値から、必要な帯域量を算出する。読み出しの場合、申告パラメータ処理部401は、算出した帯域量を受付判断部403に通知する。書き込みの場合、申告パ

ラメータ処理部401は、交換装置1からサーバユニット3への出入口の帯域として申告値を採用し、記憶装置108の処理能力の部分として、申告値に記憶装置毎に決められている一定値を乗じた数値（書き込み帯域算出部402にて計算される）を、ピーク提供速度として求める。受付判断部403は、交換装置1からサーバユニット3への入口側および出口側、ならびに記憶装置108において、GetResourceを受け付けた場合の獲得リソースの総和が一定以下となるかどうかを判断する。受け付けが可能な場合、受付判断部403は、その旨をリソース記憶部404に格納すると共に、リソース割り当て部102が管理するストリーム管理テーブルに情報を追加する。このとき、ストリーム管理テーブルの記述順序（T_{min}が小さいものから大きいものの順）を乱さない位置に情報を追加する。

【0195】（5）読み出し用リソースの確保
GetResourceメッセージを受け付けられ、かつ読み出し用のリソースが獲得されているユーザのために使用する、MS到着スロットの予約が行われる。MS到着スロットの予約は、一定時間おきに行われる。ストリーム管理テーブルに記述された順序（提供ストリームの速度が速いものから遅いものの順）で、時間区間T_{rsv}の間のMS到着スロットの予約が行われる。各T_{min}の設定値に従って、予約が行われる。図4において、参照番号803～806は、予約動作を行うタイミングを示しており、参照番号807は、時間区間T_{rsv}を示している。予約スロットの決定方法を以下に示す。

【0196】1）それぞれのユーザについて、T_{min}の時間間隔でスロットを選択し、選択したスロットのそれぞれを期待スロットとする。

2）タイミング803～806では、時間区間T_{rsv}内に、T_{min}の時間間隔で、予約の期待スロットを決定する。

3）期待スロットに予約可能であれば、予約する。

4）期待スロットに予約できなければ、期待スロットの次のスロットを参照する。参照したスロットに予約可能であれば、予約する。予約できなければ、次のスロットを参照する。これを繰り返す。

5）予約する区間を外れてしまった場合にも、上記規則に従い、次の時間区間T_{rsv}内のスロットに、予約を行う。

$$D_lim < D_net + D_disk_min + 2 \times T_min \quad \cdots (2)$$

一方、GetResourceによる申告通りとなっている場合には、スロット割り当て部103、アドレス検索部104、外部ネットワークI/O制御部106にReadBlockの到着が通知される。スロット割り当て部103では、要求された帯域で、連続的にデータブロックを提供できるように、ブロック読み出し要求送出間隔を決定する。ブロック読み出し要求送出により、M

【0197】図4において、参照番号808～820は、予約の期待スロットを示している。そして、参照番号821～833は、実際に予約を行ったスロットを示している。新たにGetResourceコマンドを受け付けたときには、T_{min}の設定値に従って、既に予約されている区間のスロット予約を直ちに行う。スロットの予約方法は、これまでに説明した方法と同一とする。すべてのユーザについて、予約動作を行った時刻のMS到着スロットで、予約されていないものについては、図5における非優先フラグをONとする。

【0198】なお、MS到着スロットの予約を、次の期待スロットが到来するタイミングが早いものから順番に行うようにしてもよい。この場合、ストリーム管理テーブルには、GetResourceを受け付けたユーザに関する情報が、次の期待スロットの到来するタイミングが早いものの順番で記述されている。MS到着スロットの予約が行われると、ストリーム管理テーブルの記述順序は更新される。

【0199】（6）ブロック読み出し動作

外部からReadBlockメッセージが到着すると、その旨が、メッセージインターフェース部101を介して申告値管理部111に通知される。ReadBlockメッセージは、以下の申告パラメータを持つ。リソース記述子

ブロックのRead間隔

ReadBlock到着からブロック送出までの遅延要求値D_{lim}

ファイル識別子

ファイル内の位置情報

【0200】申告値管理部111では、ユーザからのReadBlockコマンドによるブロックの取得要求間隔が、GetResourceによる申告通りとなっているかどうかを調べる。ReadBlockの到着時刻に、コマンドに付与された遅延要求値を加算した時刻が、ブロックの要求到着時刻となる。このブロックの要求到着時刻の間隔が、申告通りになっているかどうかを調べる。違反している場合には、違反通知を行う。また、遅延要求値が、小さ過ぎないことを確認する。すなわち、交換装置1での遅延許容値をD_{net}とし、記憶装置108での遅延限界値の最小値をD_{disk_min}として、次式（2）が成立する場合には、ユーザに対して違反通知を行う。

Sブロックを取得する間隔（MS到着間隔）は、平均間隔となるようにする。ブロック読み出し要求送出時刻に達したときに、それぞれのコマンドの遅延要求値をもとに、MS到着スロットの割り付けおよび遅延限界値算出を行う。

【0201】ただし、同一サーバユニット内宛の場合には、スロットの割り付けは行わない。ブロック読み出し

要求に付与する遅延限界値 D_disk は、次式(3)のように規定される。

$$D_disk = D_lim \quad \dots (3)$$

【0202】他のサーバユニット宛の場合には、割り付けるスロットは、以下の方法で選択する。ブロック読み

$$T_req + D_lim - 2 \times T_min < t \leq T_req + D_lim \quad \dots (4)$$

【0203】ブロック読み出し要求に付与する遅延限界値 D_disk は、割り付けたスロットの時刻を t_ar

$$D_disk = t_ar - T_req - D_net \quad \dots (5)$$

【0204】スロット割り当て部103は、要求送出部105に、遅延限界値をパラメータとした、ブロック読み出し要求の送出を指示する。スロット割り当て部103は、また、予約されたが、割り当てられなかった到着スロットをONとする。図4において、矢印834~838は、ReadBlockメッセージの到着を示している。参照番号839~843は、それぞれのReadBlockメッセージに割り当てたMS到着スロットを示している。矢印844~848は、ブロック読み出し要求の送出の指示を行うタイミングを示している。要求送出部105では、アドレス検索部104により申告された宛先に、ブロック読み出し要求を送出する。内部メッセージインターフェース部107では、ブロック読み出し要求の送信を実行する。宛先となったサーバユニット3においては、内部メッセージインターフェース部107を介して内部ネットワーク制御部110に、ブロック読み出し要求が到着する。内部ネットワーク制御部110では、パラメータとして付与された遅延限界値をもとに、読み出したMSの送出を指示する時刻を算出し、記憶する。また、記憶制御部109に対してMSの読み出しを指示する。記憶制御部109では、記憶装置108より要求されたMSの読み出しを行う。内部ネットワークI/O制御部110では、記憶装置より読み出されたMSについて、ブロック読み出し要求が到着してから遅延限界値に示された時間の後に、サーバユニットから交換装置への出力を許可する。読み出されたMSは、内部メッセージインターフェース部107を介して外部ネットワークI/O制御部106に到着する。外部ネットワークI/O制御部106では、時刻 $T_req + D_lim$ にユーザへの送出を指示する。図4において、参照番号849~853は、MSが外部ネットワークI/O制御部106に到着したタイミングを示し、参照番号854~858は、MSをユーザに送出するタイミングを示している。

【0205】リソース記述子による指定がないReadBlockメッセージについては、図5における非優先フラグがONとなっているスロットに割り当てを行い、ブロック読み出し要求を送出する。

【0206】(7)ブロック書き込み動作
外部からWriteBlockコマンドと共に到着する

出し要求を送出する時刻を T_req とする。次式

(4)を満たす時刻 T の間で、このユーザのために予約されているスロットを割り付ける。予約されているスロットが複数存在する場合には、最も早い時刻を選択する。

v、交換装置1での遅延許容値を D_net として、次式(5)で求められる。

ストリームデータは、申告値管理部111において、GetResourceにより確保した帯域を越えて到着していないかどうかチェックされる。違反している場合には、違反通知が行われる。GetResourceにより確保した帯域の範囲内に到着している場合には、スロット割り当て部103において、MSに分割され、アドレス検索部104において、宛先が決定され、ブロック書き込み要求として送出される。ここで、ブロック書き込み要求には、スロット割り当て部103にて、宛先となるサーバユニットの書き込み用スロットを必ず割り付ける。ブロック書き込み要求は、割り付けられた時刻に送出される。

【0207】(第2の実施形態)図9は、本発明の第2の実施形態に係るファイルシステムの構成を示すブロック図である。図9において、本ファイルシステムは、ファイルマネージャ901と、複数のI/Oユニット902と、複数のブロック格納ユニット905と、交換装置908とを備えている。

【0208】ファイルマネージャ901は、ファイルシステム全体を管理する。すなわち、このファイルマネージャ901は、図1に示すファイルマネージャ2と同様の機能を有する。各I/Oユニット902は、ユーザからのファイルの読み出しや、書き込みの要求を受け付けて、ユーザに対して、ファイルシステムが管理する記憶装置から読み出したデータやメッセージを送信する。各ブロック格納ユニット905は、ファイルシステムが管理するデータを記憶する記憶装置を保持し、記憶装置からのデータの読み出しや書き込みを制御する。交換装置908は、ファイルマネージャ901、各I/Oユニット902、各ブロック格納ユニット905を接続し、データやメッセージのやりとりを制御する。

【0209】本ファイルシステムが管理するデータは、ブロック格納ユニット905内に存在する記憶装置に格納されている。各I/Oユニット902におけるファイルアクセス部の構成は、図10に示すようになっている。図10に示した各ブロックの機能は、図2における同一名称のブロックの機能と同様である。各ブロック格納ユニット905におけるファイルアクセス部の構成は、図11に示すようになっている。図11に示した各ブロックの機能もまた、図2における同一名称のブロッ

クと同様の機能を有している。すなわち、I/Oユニット902およびブロック格納ユニット905は、協働して、図1のサーバユニット3と同様の機能を果たしている。

【0210】(第3の実施形態)図12は、本発明の第3の実施形態に係るファイルシステムの構成を示すブロック図である。図12において、本ファイルシステムは、メッセージインターフェース部1201と、リソース管理部1202と、読み出し/書き込み制御部1203と、記憶装置1204と、申告値管理部1205とを備えている。

【0211】メッセージインターフェース部1201は、ユーザからのメッセージの送受信を制御する。リソース管理部1202は、リソースの使用状況を管理する。読み出し/書き込み制御部1203は、記憶装置1204からの読み出しや書き込みを制御する。記憶装置1204には、データが格納されている。申告値管理部1205は、読み出し、書き込みメッセージのパラメータをチェックする。前述した第1の実施形態と同様に、図12におけるファイルシステムにおいても、GetResourceメッセージによりリソースの確保を行う。リソース管理部1202は、記憶装置1204へのアクセス可能量と、現在の使用量とを管理している。

【0212】以上、本発明をいくつかの実施形態について説明したが、本発明は上記各実施形態に限定されないのは勿論である。例えば、以下のようなことが言える。

(1) メッセージの名称は、一例を示したものであり、各実施形態で使用した名称に限定されない。また、その発行順序についても規定しない。

(2) 各メッセージのパラメータは、関連する部分の一例を挙げており、これに限定されない。例えば、予めGetResourceメッセージで確保する帯域量が決まっている場合には、Tは、パラメータとして与えず、システムで記憶していてもよい。

(3) 記憶装置108は、物理的に複数に分かれている場合もあり得る。

(4) リソース管理部202では、処理能力の最大値および使用量を単位時間当たりの読み出しデータ量であるとしたが、この方法に限らない。単位時間当たりの書き込みデータ量にする方法や、読み出し、書き込み別に最大値を定める方法も考えられる。

【0213】以上説明したように、第1～第3の実施形態では、GetResourceメッセージを受け取ったときに、ファイルシステム内部の帯域を確保することによって、ファイルの読み出し、書き込みについて、遅延品質を保証することが可能となる。そして、確保した帯域を用いて複数のファイルの読み出しまたは書き込みを実現することにより、無駄な帯域を確保する必要がなく、ファイルシステム内部のリソースを有効利用することができる。

【0214】また、書き込み時には、要求帯域に一定値を乗じたものを確保する帯域とすることにより、予め書き込み用と、読み出し用にファイルシステムの処理能力を分割しておく必要がなく、ユーザからの要求に応じて容易に比率を変えることができ、リソースを効率よく使用することができる。

【0215】さらに、ゲットリソースメッセージにより申告された遅延時間を満たす帯域と、ゲットリソースメッセージの要求に従って確保すべき帯域との内、大きい方の帯域を確保可能かを判断して、帯域の割り当てを行うことにより、複数の遅延品質をサポートすることが可能となる。

【0216】また、ブロック読み出し要求に対してMS到着スロットを割り当てることにより、ブロック格納ユニットから交換装置に入力されるブロックの個数を時間的に平滑化することができる。その結果、複数のブロック格納ユニットから読み出されて交換装置に入力されるブロックが、I/Oユニットへの入力側でぶつかって交換装置内部で廃棄されることを回避することができる。また、ブロックのトラヒックの時間による変動を少なくすることができ、交換装置を効率よく利用することが可能となる。

【0217】また、MS到着スロットの予約を、確保された帯域が大きいものから小さいものの順で行うことにより、すべての帯域の遅延品質を保証しつつ、各スロットの最大登録数を一定にして、スロットを100%利用することが可能となる。このとき、要求送出数の時間変動はなくなり、最も効率良く交換装置を利用することが可能となる。

【0218】さらに、使用しないMS到着スロットを、リソースの獲得を行っていないユーザからのデータの読み出しに割り当てることにより、さらにファイルシステムの使用効率を向上させることが可能となる。

【0219】例えばビデオサーバのように、大規模でかつ連続性を要求されるデータを扱うファイルシステムにおいては、通常、複数の記憶装置が設けられ、1つのファイルのデータが各記憶装置に分散して格納されている。このような状況下では、1つのファイルシステムに対して、帯域や記憶容量(以下、単に容量と称する)の異なる記憶装置が複数個設けられた場合、入出力帯域の制御が問題となる。従来は、各記憶装置が持つ能力(帯域、容量)の差を考慮した制御をしていなかったため、各記憶装置の能力を効率的に活用することができなかった。そこで、以下には、容量や入出力帯域の異なる複数の記憶装置の制御を容易にし、しかも要求される入出力帯域に合わせて記憶装置の使用効率を最適化できるようなディスク管理装置およびそれを用いたファイルシステムについて説明する。

【0220】(第4の実施形態)図13は、本発明の第4の実施形態に係るディスク管理装置の構成を示すプロ

ック図である。図13において、本ディスク管理装置1300は、ディスク帯域管理部1301と、ディスク容量管理部1302と、分配比率決定部1303と、帯域計算部1304と、容量計算部1305と、全ディスク帯域制御部1306と、全ディスク容量制御部1307とを備えている。

【0221】ディスク帯域管理部1301は、ファイルシステムに設けられた各記憶装置の入出力帯域を管理している。ディスク容量管理部1302は、ファイルシステムに設けられた各記憶装置の容量を管理している。分配比率決定部1303は、ファイルシステムで規定されたアルゴリズムに従って、各記憶装置にデータを分配する比率を決定する。帯域計算部1304は、ディスク帯域管理部1301で管理している各記憶装置の帯域に、分配比率決定部1303で決定された分配比率の逆数を掛けた値の中で、最も値が小さいものを、全記憶装置をまとめた場合の帯域として計算する。容量計算部1305は、ディスク容量管理部1302で管理している各記憶装置の容量に、分配比率決定部1303で決定された分配比率の逆数を掛けた値の中で、最も値が小さいものを、全記憶装置をまとめた場合の容量として計算する。全ディスク帯域制御部1306は、分配比率決定部1303と帯域計算部1304とによって決められた入出力帯域に基づき、全記憶装置の帯域制御を行う。全ディスク容量制御部1307は、分配比率決定部1303と容量計算部1305とによって決められた容量に基づき、全記憶装置の容量制御を行う。

【0222】以上のように構成されたディスク管理装置1300について、その動作を以下に説明する。まず、分配比率決定部1303で実行される帯域設定アルゴリズムの一例について説明する。

【0223】図14は、記憶装置の帯域に比例してデータを分配する比率を決定する、帯域優先レイアウトについて示している。例えば、容量と帯域とが異なる3つの記憶装置がある場合に、各記憶装置へのデータの分配を、帯域に比例して行くと、各記憶装置への分配の比率は、図14に示す分配比率の値になる。また、分配比率の逆数を各記憶装置の帯域に掛けた値を実効帯域とし、分配比率の逆数を各記憶装置の容量に掛けた値を実効容量とする。また、全記憶装置まとめた場合の帯域を、各記憶装置の実効帯域の中で最も値が小さいもの、全記憶装置まとめた場合の容量を、各記憶装置の実効容量の中で最も値が小さいものと定義する。すると、全記憶装置まとめた場合は、帯域が6Mbps、容量が6GBの記憶装置と等価と見なすことができる。

【0224】図15は、記憶装置の容量に比例してデータを分配する比率を決定する、容量優先レイアウトについて示している。例えば、容量と帯域とが異なる3つの記憶装置がある場合に、各記憶装置へのデータの分配を、容量に比例して行くと、各記憶装置への分配の比率

は、図15に示す分配比率の値になる。また、分配比率の逆数を各記憶装置の帯域に掛けた値を実効帯域とし、分配比率の逆数を各記憶装置の容量に掛けた値を実効容量とする。また、全記憶装置まとめた場合の帯域を、各記憶装置の実効帯域の中で最も値が小さいもの、全記憶装置まとめた場合の容量を、各記憶装置の実効容量の中で最も値が小さいものと定義する。すると、全記憶装置まとめた場合は、帯域が4Mbps、容量が8GBの記憶装置と等価と見なすことができる。

【0225】図16は、帯域優先レイアウトと容量優先レイアウトとを混在させてデータを分配する比率を決定する、帯域設定レイアウトについて示している。例えば、容量と帯域とが異なる3つの記憶装置がある場合に、各記憶装置への分配を、帯域優先レイアウトの比率が α 、容量優先レイアウトの比率が $(1-\alpha)$ で実行すると、各記憶装置への分配の比率は、図16に示す分配比率の値になる。また、分配比率の逆数を各記憶装置の帯域に掛けた値を実効帯域とし、分配比率の逆数を各記憶装置の容量に掛けた値を実効容量とする。また、全記憶装置まとめた場合の帯域を、各記憶装置の実効帯域の中で最も値が小さいもの、全記憶装置まとめた場合の容量を、各ディスクの実効容量の中で最も値が小さいものと定義する。このとき、全ての記憶装置の実効帯域が5Mbpsを越えるように指示された場合は、 α が $3/5$ 以上であることが必要である。今、 α を $3/5$ に設定すると、全記憶装置まとめた場合は、帯域が5Mbps、容量が20/3GBの記憶装置と等価と見なすことができる。

【0226】次に、図13に示すディスク管理装置全体の動作について説明する。予めディスク帯域管理部1301には、ファイルシステムに接続された各記憶装置の帯域が設定されている。また、予めディスク容量管理部1302には、ファイルシステムに接続された各記憶装置の容量が設定されている。ファイルシステムとして必要とされる帯域が指示されると、分配比率決定部1303は、ディスク帯域管理部1301に設定されている各記憶装置の帯域の情報と、ディスク容量管理部1302に設定されている各記憶装置の容量の情報とに基づき、上記の帯域設定アルゴリズムに従って、帯域計算部1304および容量計算部1305によって記憶装置全体としての帯域と容量を計算させる。実際にデータの入出力要求が発生すると、全ディスク帯域制御部1306は、帯域計算部1304で計算した記憶装置全体としての帯域によって、各記憶装置の入出力の帯域制御を行う。また、全ディスク容量制御部1307は、容量計算部1305で計算した記憶装置全体としての容量によって、各記憶装置の入出力の容量制御を行う。

【0227】なお、上記第4の実施形態において個々に扱っている複数の記憶装置を、論理的に1つの記憶装置と見なしたものに置き換えて処理することも可能であ

る。また、上記第4の実施形態においては、全記憶装置をまとめて処理しているが、まとめる対象を全部の記憶装置ではなく一部の記憶装置に限定しても良い。

【0228】このように、入出力帯域、容量の異なる記憶装置を組み合わせる使用する場合に、個々の記憶装置を意識することなく、複数の記憶装置をまとめて帯域、容量の管理を行うことによって、管理が容易となり、また要求される帯域に合わせて使用効率を最適化することができる。

【0229】（第5の実施形態）図17は、図13に示すディスク管理装置を用いたファイルシステムの構成を示すブロック図である。図17において、本ファイルシステムは、記憶装置1708a、1708bと、ファイル属性管理部1709と、帯域管理部1710と、ファイルオープン処理部1711と、帯域獲得部1712と、入出力制御部1713とを備えている。

【0230】ハードディスク、CD-ROM、MOD、DVD、半導体メモリ等の記憶媒体を含む記憶装置1708a、1708bには、各ファイルのデータが、それぞれ分散して格納されている。ファイル属性管理部1709は、ファイルの属性として帯域特性を管理している。帯域管理部1710は、図13に示すディスク管理装置1300の機能をその一部に含み、ファイルシステムの入出力帯域を管理している。ファイルオープン処理部1711は、データファイルのオープン処理を行う。帯域獲得部1712は、データ入出力に使用する帯域の獲得を行う。入出力制御部1713は、ファイルオープン処理部1711でオープンした複数のファイルに、帯域獲得部1712で獲得した帯域を組み合わせ、複数のファイルから任意のファイルデータを選んで入出力を行う。

【0231】実際は、上記以外にも、帯域特性以外のファイル属性を管理する機能、入出力データを送受信する機能など、種々の構成要素を必要とするが、本発明とは直接の関係がないので、その説明を省略する。

【0232】以上のように構成された第5の実施形態のファイルシステムについて、その動作を以下に説明する。まず、ディスク管理装置1300は、前述した帯域設定アルゴリズムに従って、各記憶装置へのファイルデータの分配比率を計算する。この分配比率は、レイアウトプロファイル情報中に記述され、ファイルの属性の帯域特性としてファイル属性管理部1709に格納される。

【0233】今、ビデオデータを記録した第1のファイルと、音声データを記録した第2のファイルとが存在するものとする。第1のファイルの帯域特性は、入出力に必要な帯域が6Mbpsであり、そのレイアウトプロファイルは、記憶装置1708a、1708bに均等に分配されている。第2のファイルの帯域特性は、入出力に必要な帯域が0.5Mbpsであり、そのレイアウトプ

ロファイルは、記憶装置1708a、1708bに均等に分配されている。

【0234】このとき、記憶装置1708aおよび1708bには、第1および第2のファイルのデータが、それぞれ均等に分配されて格納されている。ファイル属性管理部1709は、第1および第2のファイルの帯域特性を管理している。

【0235】ファイルオープン処理部1711は、ユーザから第1のファイルのファイルオープン要求コマンドを受け取ると、所定のオープン処理を行って、ファイル属性管理部1709から対応する帯域特性を取得する。また、ファイルオープン処理部1711は、ユーザから第2のファイルのファイルオープン要求コマンドを受け取ると、所定のオープン処理を行って、ファイル属性管理部1709から対応する帯域特性を取得する。

【0236】次に、帯域獲得部1712は、ユーザから帯域獲得要求コマンドを受け取ると、ファイルオープン処理部1711が取得した帯域特性に従って、第1および第2のファイルを同時に読み出すために必要な帯域を、帯域管理部1710から獲得する。

【0237】次に、入出力制御部1713は、ユーザから読出要求コマンドを受け取ると、ファイルオープン処理部1711がオープンした第1および第2のファイルに、帯域獲得部1712が獲得した帯域を分割して個別的に割り当て、第1のファイルのビデオデータと、第2のファイルの音声データとを、同期を取りながら記憶装置1708aおよび1708bから同時に読み出す。これによって、映像と音声との同期が取れたストリームデータが作成される。

【0238】なお、上記第5の実施形態においては、各ファイルのデータは、記憶装置1708aおよび1708bに均等に分配されているが、記憶装置の帯域、容量に応じて偏って分配することも可能である。

【0239】また、上記第5の実施形態においては、帯域特性の異なる2つファイルのデータを、2つのファイルの帯域特性を加算した帯域を用いて、同時に読み出しているが、複数のファイルのデータを、いずれか1つのファイルの帯域特性により決まる帯域（最も大きい帯域）を用いて、交互に読み出すようにしても良い。このように、複数のファイルのデータを、1つの帯域を用いて交互に読み出すことにより、AV同期などのアプリケーションの作成を容易にすることができる。

【0240】ところで、ファイルシステムの一例として、複数のクライアントに映像情報を提供する形態の映像情報提供システムでは、一般に、各クライアントにおいて、送信された情報を一時記憶するためのバッファが設けられている。各クライアントは、それぞれのバッファの容量と、受信する映像のビットレートで定まる周期で、ファイルシステム内のサーバにサービスリクエストを送信し、当該サーバは、サービスリクエストに応じ

て、所定の量の映像情報を、各クライアントに送信する。しかしながら、サーバには、一度に送信可能な情報量の上限があるため、サービスリクエストが競合する場合、映像の不連続が発生することのないよう、各クライアントからのサービスリクエストのスケジューリングを行う必要がある。

【0241】そのため、従来の映像情報提供システムにおけるサーバには、周期的に処理することが求められるサービスリクエストのスケジューリングを行うためのスケジューリング装置が設けられている。従来のスケジューリング装置としては、特公平7-104792号公報に記載されたものが知られている。この装置は、スケジュール制御プログラムが、各タスク毎でかつ優先順位毎に、優先順位対応処理待ちキューを設けるようにし、これにより、より優先順位の高い処理を先に実行し得るようにするものである。

【0242】上記従来のスケジューリング装置を用いて、優先度の高いクライアント、すなわち、サービスリクエストを送信する周期がより短いクライアントからのサービスリクエストが優先して実行されるようにすることにより、遅延が発生する回数を少なくすることができる。従って、映像の不連続が発生する頻度を低くすることができる。

【0243】しかしながら、上記従来のスケジューリング装置は、より短い周期で送信される、優先順位が上位のクライアントからのサービスリクエストを、常に優先して実行させるため、下位のクライアントからのものが、長時間、実行されない場合がある。

【0244】しかも、上記下位のクライアントからのサービスリクエストが実行されるまでの時間は、他のクライアントからのサービスリクエストの個数により変動する。このため、上記下位のクライアント側で、遅延時間を考慮に入れた周期でサービスリクエストの送信を行うようにして、映像の不連続が発生するのを防ぐことができない。

【0245】また、全クライアントから送信されるサービスリクエストの個数の時間的な平均が、サーバの処理能力を超えている場合には、時間の経過とともに、処理しきれずに残されるサービスリクエストの数が増加する。この場合、当然ながら、どのような順序でサービスリクエストを実行させても、映像の不連続が発生するのを防ぐことができない。

【0246】また、複数のサーバが、共通の伝送路を介して、クライアントに映像情報を提供する形態の映像情報提供システムでは、伝送路に、一度に伝送可能な情報量の上限があるため、各サーバへのサービスリクエストのスケジューリングを行う必要がある。そこで、上記従来のスケジューリング装置を用いてスケジューリングを行った場合、上記と同様の課題が生じる。

【0247】そこで、以下には、サーバから複数のクラ

イアントに、映像情報などの連続した情報が分割して提供される際、情報の不連続が発生しないよう、各クライアントからのサービスリクエストのスケジューリングを行うことができるスケジューリング装置について説明する。

【0248】また、複数のサーバからクライアントに、連続した情報が分割して提供される際、情報の不連続が発生しないよう、各サーバへのサービスリクエストのスケジューリングを行うことができるスケジューリング装置について説明する。

【0249】（第6の実施形態）以下、本発明の第6の実施形態について、図面を参照しながら説明する。図18は、本発明の第6の実施形態に係るスケジューリング装置の構成を示すブロック図である。図18の装置は、登録指示部1811、実行権発生周期テーブル1812、実行権発生部1813、処理待ちキュー1814および実行指示部1815を備えている。実行指示部1815は、実行権キュー1816を含む。

【0250】図19は、図18のスケジューリング装置を用いた、映像情報提供システムの構成を示すブロック図である。図19のシステムは、ファイルシステムの一例の映像情報提供装置1920および複数のクライアント(1~n)1921で構成される。映像情報提供装置1920は、スケジューリング装置1810およびサーバ1922を含む。

【0251】図19のクライアント(1~n)1921はそれぞれ、映像情報提供装置1920のサーバ1922に、周期的に読み出し要求を送信する。クライアント(1~n)1921にはバッファおよびモニタが設けられており、クライアント(1~n)1921は、提供される映像情報をこのバッファに一時蓄え、蓄えた情報を読み出ししながら、モニタに映像を表示する。スケジューリング装置1810は、サーバ1922から、クライアント(1~n)1921に、それぞれ周期的に映像情報が提供されるよう、読み出し要求のスケジューリングを行う。サーバ1922は、スケジューリング装置1810の指示に応じて、クライアント(1~n)1921に、映像情報を所定の量ずつ送信する。

【0252】図18の登録指示部1811は、図19のクライアント(1~n)1921からの申告を受けて、クライアント(1~n)1921毎の、実行権を発生させるべき周期を、実行権発生周期テーブル1812に指示して登録させる。実行権発生周期テーブル1812は、実行権を発生させるべき周期を登録する。実行権発生部1813は、所定の周期(以下、 T_{s1t})で起動され、実行権発生周期テーブル1812に登録された周期に基づいて、クライアント(1~n)1921毎の実行権を発生させる。実行権発生部1813はまた、発生させた実行権に、優先度を示す評価指数(以下、 $N_{sf t}$)を付与する。処理待ちキュー1814は、クライ

ント(1~n)1921からの読み出し要求をキューイングする。実行指示部1815は、周期Tsl tで起動され、処理待ちキュー1814にキューイングされている読み出し要求のうち、実行権発生部1813が発生させた実行権に対応するクライアント1921からの読み出し要求を取得して、その読み出し要求を実行するよう、サーバ1922に指示する。

【0253】図20は、図19の映像情報提供装置1920とクライアント(1~n)1921との間で、信号が送受信される様子を示すシーケンス図である。図20のリソース獲得要求31は、送信元のクライアント1921を識別するための情報(以下、クライアントID)と、読み出しを希望する映像情報名および読み出し位置とを通知するために、クライアント(1~n)1921が映像情報提供装置1920に送信する信号である。リソース獲得要求31には、さらに、クライアント(1~n)1921がそれぞれ希望する、読み出し要求を送信する周期、すなわち、上記実行権を発生させるべき周期(以下、Treq)が含まれる。リソース獲得応答32は、希望されたTreqで、映像情報を提供できるか否かを通知するために、映像情報提供装置1920がクライアント(1~n)1921に送信する信号である。

【0254】読み出し要求33は、読み出しを希望する際に、クライアント(1~n)1921が映像情報提供装置1920に送信する信号である。読み出し応答34は、読み出し要求33に応じて、映像情報提供装置1920がクライアント(1~n)1921に送信する映像情報である。リソース開放要求35は、希望する映像情報の読み出しが完了したことを通知するために、クライアント(1~n)1921が映像情報提供装置1920に送信する信号であって、クライアントIDが含まれる。リソース解放応答36は、リソース開放要求35に応じて、映像情報の提供を停止することを通知するために、映像情報提供装置1920がクライアント(1~n)1921に送信する信号である。

【0255】図21は、図18の実行権発生周期テーブル1812の登録内容を示す図である。図21に示すように、実行権発生周期テーブル1812は、クライアント(1~n)1921毎のクライアントID、Treqおよび実行権発生周期オフセット(以下、Tofs)を登録する。図22は、図18の処理待ちキュー1814のキューイング内容を示す図である。図22に示すように、処理待ちキュー1814は、クライアント(1~n)1921からの読み出し要求を、クライアント(1~n)1921毎に、キューイングする。

【0256】図23は、図18の実行権キュー1816のキューイング内容を示す図である。図23に示すように、実行権キュー1816は、発生された実行権を、付与されたNsft毎に、キューイングする。図24は、図18の実行権発生部1813が、実行権を発生させる

動作を説明するためフローチャートである。

【0257】図25は、図18の実行指示部1815が、図19のサーバ1922に、読み出し要求の実行を指示する動作を説明するためのフローチャートである。

【0258】以下には、図19のスケジューリング装置1810が、サーバ1922から、複数のクライアント(1~n)1921に、それぞれ周期的に映像情報が提供されるよう、読み出し要求のスケジューリングを行う動作について、図18~25を参照しながら説明する。最初、図18の登録指示部1811が、実行権発生周期テーブル1812に指示して、クライアントID、TreqおよびTofsを登録させる動作について説明する。

【0259】いま、実行権発生周期テーブル1812には、複数のクライアント1921のクライアントID、TreqおよびTofsが、すでに登録されているものとする。この時、他のクライアント1921が、図20のリソース獲得要求31を送信することにより、自己のクライアントIDおよびTreqを映像情報提供装置1920に申告する。

【0260】登録指示部1811は、サーバ1922の情報提供能力、すなわち、周期Tsl tに等しい時間内にサーバ1922が実行可能な読み出し要求の最大数

(以下、Nmax)を認識しており、すでに登録されているものに加えて、申告されたTreqが登録された場合、発生される実行権のTsl t当たりの平均数と、Nmaxとを比較することにより、サーバ1922が、時間的に平均して、情報を提供できるか否かを判定する。

【0261】情報を提供できると判定した場合、登録指示部1811は、申告されたクライアントIDおよびTreqを、実行権発生周期テーブル1812に指示して登録させ、また、Tofsの初期値として、“0”を登録させる。そして、登録指示部1811は、リソース獲得応答32を送信することにより、申告されたTreqで映像情報を提供できることを、リソース獲得要求31の送信元のクライアント1921に通知する。

【0262】情報を提供できないと判定した場合には、登録指示部1811は、リソース獲得応答32を送信することにより、申告されたTreqで情報を提供できないことを、送信元のクライアント1921に通知する。

【0263】次に、図18の実行権発生部1813が、実行権発生周期テーブル1812に登録されたTreqおよびTofsに基づいて、実行権を発生させる動作について、図24を参照しながら説明する。なお、実行権発生周期テーブル1812には、複数のクライアント1921のクライアントID、TreqおよびTofsが、上記のようにして登録されているものとする。

【0264】実行権発生部1813は、最初、実行権発生周期テーブル1812に登録されたクライアントID、TreqおよびTofsを一組取り出す(ステップ

S1)。次に、実行権発生部1813は、取り出したTofsと、自己の動作の周期Ts1tとを比較する(ステップS2)。Tofs \geq Ts1tである場合、実行権発生部1813は、ステップS9に進む。

【0265】Tofs<Ts1tである場合には、実行権発生部1813は、読み出したクライアントIDに対応するクライアント1921からの読み出し要求が、処理待ちキュー1814にキューイングされているか否かを判定する(ステップS3)。キューイングされている場合、実行権発生部1813は、実行権を発生させるとともに、その読み出し要求に実行権が割り当てられたことを示す情報を付与する一方、次式(6)により、新たなTofsを算定する(ステップS4)。

【0266】 $Tofs = Tofs + Treq \dots (6)$
キューイングされていない場合には、実行権発生部1813は、ステップS9に進む。

【0267】次に、実行権発生部1813は、算定して得た新たなTofsと、Ts1tとを比較する(ステップS5)。Tofs<Ts1tである場合、実行権発生部1813は、発生させた実行権に、評価指数として、Nsft=0を付与する(ステップS6)。発生された実行権を、実行指示部1815が、実行権キュー1816に指示して、Nsftの値が一致するキューの最後尾にキューイングさせると(ステップS7)、実行権発生部1813は、ステップS2に戻る。

【0268】Tofs \geq Ts1tである場合には、実行権発生部1813は、発生させた実行権に、評価指数として、次式(7)により算定されるNsftを付与し(ステップS8)、ステップS7に進む。

$Nsft = f(Treq / Ts1t) \dots (7)$

ただし、上式(7)において、f(x)はx以上の最小の整数である。

【0269】ステップS2で、Tofs \geq Ts1tである場合、実行権発生部1813は、次式(8)により、新たなTofsを算定する(ステップS9)。

$Tofs = Tofs - Ts1t \dots (8)$

【0270】次に、実行権発生部1813は、算定して得た新たなTofsが、Tofs \geq 0であるか否かを判定する(ステップS10)。Tofs \geq 0である場合、実行権発生部1813は、実行権発生周期テーブル1812に指示して、Tofsを、算定して得た新たなTofsに更新させる(ステップS11)。

【0271】Tofs<0である場合には、実行権発生部1813は、Tofsを、0に更新させる(ステップS12)。

【0272】次に、実行権発生部1813は、実行権発生周期テーブル1812に登録されている全てのクライアントID、TreqおよびTofsを取り出したか否かを判定する(ステップS13)。全て取り出したと判定した場合、実行権発生部1813は、動作を停止し、

取り出すべきものがあると判定した場合には、ステップS1に戻る。

【0273】次に、図18の実行指示部1815が、上記のようにして発生された実行権に基づいて、サーバ1922に対して、キューイングされている読み出し要求の実行を指示する動作について、図25を参照しながら説明する。実行指示部1815は、サーバ1922に実行を指示した回数をカウントしており、カウントして得たカウント値Cを記憶している。実行指示部1815は、最初、このカウント値Cを初期化する(ステップS21)。

【0274】次に、実行指示部1815は、実行権キュー1816にキューイングされている実行権のうち、Nsftの値が最小のキューにキューイングされている実行権の先頭のもの、すなわち、付与されたNsftが最小で、かつ、発生されてからの経過時間が最も長いものを取り出し(ステップS22)、次いで、カウント値Cをインクリメントする(ステップS23)。

【0275】次に、実行指示部1815は、取り出した実行権に付与されたクライアントIDを参照して、処理待ちキュー1814から、その実行権に対応するクライアント1921からの読み出し要求を取り出して、サーバ1922に実行を指示する(ステップS24)。そして、実行指示部1815は、インクリメントして得たカウント値Cを、Nmaxと比較する(ステップS25)。

【0276】C<Nmaxである場合、実行指示部1815は、ステップS22に戻り、C \geq Nmaxである場合には、実行権キュー1816に指示して、残された実行権のNsftの値をそれぞれデクリメントさせ(ステップS26)、その後、処理を終了する。

【0277】上記のようにして実行の指示を受けたサーバ1922は、指示を受けてから、ある一定の時間が経過した後、指定された映像情報を、読み出し要求の送信元であるクライアント1921に送信する。クライアント1921は、必要な映像情報の受け取りを完了すると、図20のリソース解放要求35を送信することにより、受け取りが完了した意を、映像情報提供装置1920に通知する。リソース解放要求35には、送信元のクライアント1921のクライアントIDが付与されており、リソース解放要求35を受信した登録指示部1811は、実行権発生周期テーブル1812から、付与されたクライアントIDと、そのクライアントIDに対応するTreqおよびTofsとを削除する。

【0278】上記のように、本実施形態によれば、サーバ1922から、複数のクライアント(1~n)1921に、それぞれ周期的に映像情報が提供されるよう、読み出し要求のスケジューリングを行う際、処理待ちキュー1814は、各クライアント(1~n)1921からの読み出し要求をキューイングする。実行権発生周期テ

ーブル1812は、クライアント(1~n)1921毎の、実行権を発生させるべき周期 $Treq$ を登録し、実行権発生部1813は、実行権発生周期テーブル1812に登録された周期に基づいて、実行権を発生させる。実行指示部1815は、処理待ちキュー1814にキューイングされている読み出し要求のうち、実行権発生部1813が発生させた実行権に対応するクライアント1921からの読み出し要求を取得して、サーバ1922に実行を指示する。

【0279】このように、各クライアント(1~n)1921からの読み出し要求をキューイングする一方、クライアント(1~n)1921毎の、実行権を発生させるべき周期を登録しておいて、その周期に基づいて、実行権を発生させる。そして、キューイングしている読み出し要求のうち、発生させた実行権に対応するクライアント1921からのものを、サーバ1922に指示して実行させる。これにより、あるクライアント1921からの読み出し要求が実行されるまでの時間が、他のクライアント1921からの読み出し要求の送信の状況に影響されることがなくなり、従って、送信された読み出し要求が実行されるまでの時間が不定となるのを防ぐことができる。

【0280】また、実行権発生部1813は、周期 Tsl で起動され、実行権発生周期テーブル1812を参照して、起動された現時刻から、次の実行権を発生させる時刻までの時間 $Tofs$ が、 $Tofs < Tsl$ であるクライアント1921の実行権を発生させる。さらに、当該クライアント1921の $Tofs$ に $Treq$ を加算し、加算して得た $Tofs$ が、 $Tofs \geq Tsl$ である場合、当該 $Tofs$ に対して、式 $Tofs = Tofs - Tsl$ により、新たな $Tofs$ を算定し、算定して得た $Tofs$ を、実行権発生周期テーブル1812に指示して登録させる。

【0281】加算して得た $Tofs$ が、 $Tofs < Tsl$ である場合には、実行権発生部1813は、さらに、実行権を発生させるとともに、 $Tofs$ に $Treq$ を加算して Tsl と比較する動作を繰り返す。そして、加算して得た $Tofs$ が、 $Tofs \geq Tsl$ となった時点で、その $Tofs$ に対して、式 $Tofs = Tofs - Tsl$ により、新たな $Tofs$ を算定し、算定して得た $Tofs$ を、実行権発生周期テーブル1812に指示して登録させる。

【0282】なお、起動された現時刻から、次の実行権を発生させる時刻までの時間 $Tofs$ が、 $Tofs \geq Tsl$ であるか、または、読み出し要求が、処理待ちキュー1814にキューイングされていないクライアント1921については、実行権を発生させず、その $Tofs$ に対して、式 $Tofs = Tofs - Tsl$ により、新たな $Tofs$ を算定し、算定して得た $Tofs$ を、実行権発生周期テーブル1812に指示して登録させる。

【0283】算定して得た $Tofs$ が、 $Tofs < 0$ である場合には、実行権発生部1813は、 $Tofs = 0$ を、実行権発生周期テーブル1812に指示して登録させる。

【0284】このように、実行権発生部1813は、周期 Tsl で起動され、起動されると、実行権発生周期テーブル1812に登録された $Treq$ と、 Tsl とに基づいて、 $Tofs$ 、すなわち、起動された現時刻から、次の実行権を発生させる時刻までの時間を算定し、算定して得た $Tofs$ と、 Tsl とを比較することにより、実行権を発生させるべきクライアント1921を判定して、実行権を発生させる。さらに、実行権発生部1813は、 $Tofs$ に $Treq$ を加算して、 Tsl と比較する動作を繰り返すことにより、複数の実行権を発生させべきクライアント1921を判定して、必要な数の実行権を発生させる。実行権発生部1813は、これらの動作を、起動時に一括して行うことにより、実行権発生周期テーブル1812に登録された周期 $Treq$ で実行権を発生させるのに比べて、実行権発生部1813の処理動作を軽減できる。

【0285】また、読み出し要求がキューイングされていないクライアント1921については、実行権を発生させないことにより、読み出し要求を送信しなかったクライアント1921に、映像情報が提供されないようにすることができる。

【0286】また、算定して得た新たな $Tofs$ が、 $Tofs < 0$ である場合には、 $Tofs = 0$ を登録させることにより、読み出し要求がキューイングされていないクライアント1921の $Tofs$ が、繰り返し Tsl を減算されて小さな値となり、読み出し要求がキューイングされた時点で、そのクライアント1921の実行権が多数発生されるのを防ぐことができる。

【0287】さらに、実行権発生部1813は、上記のようにして発生させた実行権に、上記の式(7)により算定される評価指数 $Nsft$ を付与する。実行指示部1815は、実行権発生部1813が発生させた実行権をキューイングする実行権キュー1816を含んでおり、発生された実行権を、このキューに指示して、評価指数毎でかつ発生された順にキューイングさせる。そして、実行指示部1815は、キューイングされている実行権のうち、付与された評価指数 $Nsft$ がより小さく、かつ、より先に発生された実行権を選択して、その実行権に対応するクライアント1921からの読み出し要求の実行を、サーバ1922に指示する。

【0288】実行指示部1815は、周期 Tsl に等しい時間内にサーバ1922が実行可能な読み出し要求の最大数 $Nmax$ を認識しており、この $Nmax$ に等しい個数だけ、実行を指示した時点で、実行の指示を停止し、かつ、実行権キュー1816に指示して、残された実行権の評価指数 $Nsft$ をディクリメントさせる。

【0289】このように、実行権発生部1813が、発生させた実行権に、実行権を発生させるべき周期の関数として算定した、優先度を示す評価指数を付与し、実行指示部1815は、付与された評価指数に基づいて、優先されるべき実行権を選択して、実行を指示する。そして、実行指示部1815は、 N_{max} 個の読み出し要求の実行を指示した時点で、実行の指示をいったん停止し、さらに、残された実行権については、その評価指数をディクリメントして、優先度をより高くする。

【0290】これにより、実行指示部1815は、起動された現時刻から、次の実行権が発生される時刻までの時間が、より短いクライアント1921の実行権を、優先して選択することができ、従って、遅延が発生する回数をより少なくできる。また、処理能力を超える数の読み出し要求を、サーバ1922に実行させるのを防ぐことができる。とともに、残された実行権については、評価指数が T_{slt} 毎にディクリメントされるため、発生されてから、遅くとも $T_{req} + T_{slt}$ 後までに、当該実行権が選択されるようにすることができる。従って、クライアント(1~n)1921は、この最大遅延時間を考慮に入れた周期で、読み出し要求を送信することにより、映像情報の不連続が発生するのを防ぐことができる。

【0291】ただし、時間的に平均して、サーバ1922の処理能力を超える数の実行権が発生される場合には、時間の経過に従って、残される実行権の数が増加する。この場合には、上記のような処理を行っても、映像の不連続が発生するのを防ぐことができない。

【0292】そこで、登録指示部1811は、 T_{req} を登録するよう申告を受けると、すでに登録されているものに加えて、申告された T_{req} がさらに登録されても、発生される実行権の T_{slt} 当たりの平均数が、 N_{max} を越えないと判定した場合にのみ、実行権発生周期テーブル1812に指示して、申告された T_{req} を登録させる。

【0293】これにより、時間的に平均して、発生される実行権の総数がサーバ1922の処理能力を越えているために、時間の経過とともに、残される実行権の数が増加して、映像の不連続が発生するのを防ぐことができる。

【0294】また、実行指示部1815に実行権キュー1816を設けて、評価指数毎でかつ発生された順にキューイングさせることにより、実行指示部1815は、実行権を選択する際、評価指数が最小のキューの先頭から順に選択すればよい。ため、実行指示部1815の処理動作を軽減できる。

【0295】(第7の実施形態)以下、本発明の第7の実施形態について、図面を参照しながら説明する。図26は、本発明の第7の実施形態に係るスケジューリング装置の構成を示すブロック図である。図26の装置は、

登録指示部2691、実行権発生周期テーブル2692、実行権発生部2693、処理待ちキュー2694および実行指示部2695を備えている。実行指示部2695は、実行権キュー2696を含む。

【0296】図27は、図26のスケジューリング装置を用いた、映像情報提供システムの構成を示すブロック図である。図27のシステムは、映像情報提供装置2700およびクライアント2701で構成され、映像情報提供装置2700とクライアント2701との間は、伝送路2703で接続されている。映像情報提供装置2700は、スケジューリング装置2690および複数のサーバ(1~m)2702を含む。

【0297】図27のクライアント2701は、映像情報提供装置2700の複数のサーバ(1~m)2702に、読み出し要求をそれぞれ周期的に送信する。クライアント2701には、複数のバッファおよびモニタが設けられており、クライアント2701は、提供される情報を、サーバ(1~m)2702毎に、各バッファに一時的に蓄え、蓄えた情報を読み出しながら、各モニタに映像を表示する。スケジューリング装置2690は、サーバ(1~m)2702から、クライアント2701に、それぞれ周期的に映像情報が提供されるよう、読み出し要求のスケジューリングを行う。サーバ(1~m)2702はそれぞれ、スケジューリング装置2690の指示に応じて、クライアント2701に、映像情報を所定の量ずつ送信する。なお、映像情報提供装置2700に含まれるサーバ(1~m)2702の情報提供能力の合計は、伝送路2703の伝送能力を上回っているものとする。

【0298】図26の登録指示部2691は、図27のクライアント2701からの申告を受けて、サーバ(1~m)2702毎の、実行権を発生させるべき周期を、実行権発生周期テーブル2692に指示して登録させる。実行権発生周期テーブル2692は、実行権を発生させるべき周期を登録する。実行権発生部2693は、所定の周期(以下、 T_{slt})で起動され、実行権発生周期テーブル2692に登録された周期に基づいて、サーバ(1~m)2702毎の実行権を発生させる。実行権発生部2693はまた、発生させた実行権に、優先度を示す評価指数(以下、 N_{sft})を付与する。処理待ちキュー2694は、サーバ(1~m)2702への読み出し要求をキューイングする。実行指示部2695は、周期 T_{slt} で起動され、処理待ちキュー2694にキューイングされている読み出し要求のうち、実行権発生部2693が発生させた実行権に対応するサーバ2702への読み出し要求を取得して、その読み出し要求を実行するよう、対応するサーバ2702に指示する。

【0299】また、第6の実施形態において図20を用いて説明したものと同様の信号が、図27の映像情報提供装置2700とクライアント2701との間で送受信

される。ただし、リソース獲得要求31には、クライアントIDに代えて、サーバIDが含まれる。

【0300】図28は、図26の実行権発生周期テーブル2692の登録内容を示す図である。図28に示すように、実行権発生周期テーブル2692は、サーバ(1~m)2702毎のサーバID、T i t vおよび実行権発生周期オフセット(以下、T o f s)を登録する。なお、図21のテーブルのT r e qは、図19のクライアント(1~n)1921から申告された値であるが、図28のT i t vは、図27のクライアント2701から申告されたT r e qに基づいて、所定の方法で算定された値である。

【0301】図29は、図26の処理待ちキュー2694のキューイング内容を示す図である。図29に示すように、処理待ちキュー2694は、サーバ(1~m)2702への読み出し要求を、サーバ(1~m)2702毎に、キューイングする。

【0302】図30は、図26の実行権キュー2696のキューイング内容を示す図である。図30に示すように、実行権キュー2696は、発生された実行権を、付与されたN s f t毎に、キューイングする。

【0303】以下には、図27のスケジューリング装置2690が、複数のサーバ(1~m)2702から、クライアント2701へ、伝送路2703を介して、それぞれ周期的に映像情報が提供されるよう、読み出し要求のスケジューリングを行う動作について、図20、24~30を参照しながら説明する。最初、図26の登録指示部2691が、実行権発生周期テーブル2692に指示して、サーバID、T i t vおよびT o f sを登録さ

$$T i t v = 1 / (1 / T i t v + 1 / T r e q) \quad \cdots (9)$$

【0307】また、登録指示部2691は、実行権発生周期テーブル2692に指示して、T o f sを、下記の

$$T o f s = T o f s + (T i t v (n e w) - T i t v (o l d)) \quad \cdots (10)$$

ただし、上式(10)において、T i t v (n e w)は、更新されたT i t vを、T i t v (o l d)は更新される前のT i t vを、それぞれ意味する。

【0308】少なくとも一方の能力がないと判定した場合には、登録指示部2691は、クライアント2701に対して、リソース獲得応答32を送信することにより、要求を受け入れないことを通知する。

【0309】なお、上記(a)および(b)について、両方の能力があると判定し、かつ指定されたサーバ2702のサーバIDが実行権発生周期テーブル2692に登録されていない場合には、登録指示部2691は、実行権発生周期テーブル2692に指示して、サーバID、T i t vおよびT o f sを新たに登録させる。この場合、T i t vとして、申告されたT r e qが、T o f sとして、“0”が、それぞれ登録される。

【0310】実行権発生部2693が、実行権発生周期

せる動作について説明する。

【0304】いま、実行権発生周期テーブル2692には、複数のサーバ2702のサーバID、T i t vおよびT o f sが、すでに登録されているものとする。この時、クライアント2701が、図20のリソース獲得要求31を送信することにより、希望するサーバ2702のサーバIDおよびT r e qを、映像情報提供装置2700に申告する。

【0305】登録指示部2691は、サーバ(1~m)2702の情報提供能力および伝送路2703の伝送能力を認識しており、下記の(a)および(b)について判定する。

(a) 指定されたサーバ2702の情報提供能力が、すでに登録されているT i t vを、申告されたT r e qを含めた新たなT i t vに更新した場合、情報を提供できるものであるか。

(b) 伝送路2703の伝送能力が、すでに登録されているT i t vを、申告されたT r e qを含めた新たなT i t vに更新した場合、サーバ(1~m)2702から提供される情報を伝送できるものであるか。

【0306】上記(a)および(b)について、両方の能力があると判定した場合、登録指示部2691は、実行権発生周期テーブル2692に指示して、すでに登録されている、指定されたサーバ2702のT i t vを、下記の式(9)で算定される新たなT i t vに更新させるとともに、クライアント2701に対して、図20のリソース獲得応答32を送信することにより、要求を受け入れることを通知する。

式(10)により算定される新たなT o f sに更新させる。

テーブル2692に登録されたT i t vおよびT o f sに基づいて実行権を発生させる動作は、第6の実施形態において、図24を用いて説明したものと同様である。ただし、実行権発生周期テーブル2692には、複数のサーバ(1~m)2702のサーバID、T i t vおよびT o f sが、上記のようにして登録されている点異なる。

【0311】また、図24のステップS4において、式(6)に代えて、下記の式(11)により、新たなT o f sが算定される点も異なる。

$$T o f s = T o f s + T i t v \quad \cdots (11)$$

【0312】同様に、ステップS8において、式(7)に代えて、下記の式(12)により、N s f tが算定される点も異なる。

$$N s f t = f (T i t v / T s l t) \quad \cdots (12)$$

ただし、上式(12)において、 $f(x)$ は x 以上の最小の整数である。

【0314】次に、実行指示部2695は、付与された評価指数に基づいて実行権を選択し、処理待ちキュー2694にキューイングされている読み出し要求のうち、選択した実行権に対応するサーバ2702への読み出し要求を取得して、対応するサーバ2702に実行を指示する。その際の実行指示部2695の動作は、第6の実施形態において、図25を用いて説明したものと同様である。ただし、カウント値 C は、実行指示部2695が全サーバ(1~ m)2702に実行を指示した合計回数であり、 N_{max} は、周期 T_{slit} に等しい時間内に伝送路2703が伝送可能な最大の情報量に対応する読み

$$T_{itv} = 1 / (1 / T_{itv} - 1 / T_{req}) \quad \dots (13)$$

また、登録指示部2691は、実行権発生周期テーブル2692に指示して、 T_{ofs} を、上記の式(10)により算定される新たな T_{ofs} に更新させる。

【0317】上記のように、本実施形態によれば、複数のサーバ(1~ m)2702から、クライアント2701に、それぞれ周期的に映像情報が提供されるよう、読み出し要求のスケジューリングを行う際、第6の実施形態におけるものと同様の動作を行う。さらに、登録指示部2691は、すでに登録されている、あるサーバ2702の実行権を発生させるべき周期 T_{itv} を、さらに周期 T_{req} で実行権を発生させるものに更新するよう申告を受けた際、すでに登録されている T_{itv} を、申告された T_{req} を含めた新たな T_{itv} に更新しても、伝送路2703の伝送能力が、全サーバ(1~ m)2702から提供される情報を伝送できるものである場合にのみ、実行権発生周期テーブル2692に指示して、登録されている T_{itv} を、上記の式(9)により算定される新たな周期 T_{itv} に更新させる。

【0318】これにより、登録された周期を更新した際、発生される実行権の T_{slit} 当たりの平均数が、伝送路の伝送能力を越えているために、時間の経過ともなって、残される実行権の数が増加するのを防ぐことができる。

【0319】また、登録指示部2691は、すで登録されている、あるサーバ2702の実行権を発生させるべき周期 T_{itv} を、周期 T_{req} を除いて実行権を発生させるものに更新するよう申告を受けた際、実行権発生周期テーブル2692に指示して、登録されている T_{itv} を、上記の式(13)により算定される新たな周期 T_{itv} に更新させる。

【0320】これにより、すでに登録されている、実行権を発生させるべき周期を、周期 T_{req} で発生させている実行権を除いて発生させるものに更新するよう申告を受けた場合、すでに登録されている周期を、新たなものに更新することができる。

【0321】なお、本実施形態では、複数のバッファお

出し要求の個数である。

【0315】上記のようにして映像情報が提供され、提供が完了すると、クライアント2701は、第6の実施形態においてクライアント1921が行ったのと同様に、図20のリソース解放要求35を送信することにより、希望する映像情報の受け取りが完了したことを映像情報提供装置2700に通知する。ただし、クライアント2701が送信するリソース解放要求35には、サーバIDおよび T_{req} が付加されており、図26の登録指示部2691は、実行権発生周期テーブル2692に指示して、登録されている T_{itv} を、下記の式(13)により算定される新たな T_{itv} に更新させる。

【0316】

よびモニタを備えたクライアント2701を、一つだけ含むようにしているが、共通の伝送路2703を介して、一つのバッファおよびモニタを備えたクライアント2701を、複数含むようにしてもよい。ただし、この場合、各クライアント2701は、それぞれに定められた周期で、読み出し要求を送信する必要がある。

【0322】また、本実施形態では、伝送路2703の伝送能力の制限があるために、スケジューリングを行うとしているが、クライアント2701の受信能力(バッファの容量や処理速度など)の制限があるために、スケジューリングを行うとしてもよい。

【0323】また、第6および第7の実施形態において、実行権発生部1813および2693ならびに実行指示部1815および2695は、ともに周期 T_{slit} で起動されるとしているが、それぞれ互いに異なる周期で起動されるようにしてもよい。

【0324】また、第6の実施形態において、 N_{max} は、周期 T_{slit} に等しい時間内にサーバ1922が実行可能な読み出し要求の最大数であるとしているが、単位時間にサーバ1922が実行可能な読み出し要求の最大数であるとしてもよい。同様に、第7の実施形態において、 N_{max} は、周期 T_{slit} に等しい時間内に伝送路2703が伝送可能な情報量に対応する読み出し要求の個数であるとしているが、単位時間に伝送路2703が伝送可能な情報量に対応する読み出し要求の個数であるとしてもよい。

【0325】また、図18のスケジューリング装置と、図26のスケジューリング装置とを含む他のスケジューリング装置を用いることにより、複数のサーバから、複数のクライアントに、それぞれ周期的に映像情報が提供されるよう、スケジューリングを行うことができる。すなわち、最初、図18のスケジューリング装置が、クライアント毎の実行権を発生させ、発生させた実行権のうち、あるクライアントの実行権を選択して、その実行権に対応する読み出し要求を、図26のスケジューリング装置に入力する。次に、図26のスケジューリング装置

が、サーバ毎の実行権を発生させ、発生させた実行権のうち、あるサーバの実行権を選択して、入力された読み出し要求を実行するよう、そのサーバに指示する。

【0326】また、第6および第7の実施形態において、実行権発生部1813および2693はそれぞれ、上記の式(7)、式(12)により、Nsftを算定するとしたが、TreqをTslitで除して得られる値の整数部分をNsftとするとしてもよい。

【0327】また、実行権発生周期テーブル1812および2692が、実行権発生周期オフセットTofsに代えて、次の実行権を発生させる時間を示すTnextを登録し、実行権発生部1813および2693は、現時間がTnextに一致した時点で、実行権を発生させるようにしてもよい。

【0328】また、実行権発生部1813および2693がNsftを算定し、実行指示部1815および2695は、Nsftがより小さい実行権を選択するとしたが、実行権発生部1813および2693が上記のTnextを算定し、実行指示部1815および2695は、Tnextが現時間により近い実行権を選択するとしてもよい。

【0329】また、実行指示部1815および2695は、カウント値Cが、 $C \geq N_{max}$ となった場合、それぞれ実行権キュー1816、実行権キュー2696に指示して、残された実行権のNsftをディクリメントさせるとしたが、実行権キュー1816および2696の代わりに、図31に示すような他の実行権キューを用い、実行指示部1815および2695は、カウント値Cが、 $C \geq N_{max}$ となった場合、上記他の実行権キューに指示して、ポインタをずらすようにしてもよい。なお、図31に示すように、上記他の実行権キューはリング構造を有しており、ポインタの位置が移動するのにともなって、Nsftの値が順次入れ替わる。

【0330】前述したように、CATVネットワークにおけるビデオオンデマンドシステムや放送局の映像編集システム等のファイルシステムにおいては、映像や音声等のデータを扱うサーバ装置が設けられている。このサーバ装置は、ユーザ端末あるいはクライアントからデータの読み出し要求を受けたとき、内部の記憶装置を検索して要求に対応するデータを読み出し、読み出しを要求したユーザ端末あるいはクライアントに、当該データを送出する機能を有している。

【0331】ところで、従来のサーバ装置は、送出するデータの連続性を保証するために、ユーザ端末へデータを送出する前に、送出するデータの量に見合う、サーバ装置内部のリソース（単位時間当たりにデータを処理または転送できる能力）の使用権と、サーバ装置とデータの取得を要求するユーザ端末との間の通信リソース（単位時間当たりにデータを送信できる能力）の使用権とを確保した上で、データの送出を実行するようになってい

る。

【0332】しかしながら、上記のような従来のサーバ装置では、動画や音声のような連続性を必要とするデータのみならず、静止画のような連続性を必要としないデータの送出に関しても、データの送出前にリソース使用権を確保するようにしている。そのため、連続性を必要としないデータの送出量が増えるに従い、連続性を必要とするデータのリソース確保が困難になり、最悪の場合はデータの送出が行えなくなるという問題が生じていた。

【0333】そこで、以下には、連続性を必要としないデータを送出する際に効率的にリソースを使用でき、しかも、連続性を必要としないデータの送出量が増加しても、連続性を必要とするデータを送出するためのリソース確保に影響を与えないサーバ装置について説明する。

【0334】（第8の実施形態）図32は、本発明の第8の実施形態に係るサーバ装置を用いたデータ伝送システムの構成を示すブロック図である。図32において、本データ伝送システムは、リソース管理部3201と、データ記憶部3202と、交換部3203と、入出力部3204と、複数のユーザ端末3208とを備えている。なお、本実施形態のサーバ装置は、リソース管理部3201と、データ記憶部3202と、交換部3203と、入出力部3204とによって構成される。そして、当該サーバ装置と各端末3208は、有線または無線の伝送路を介して接続されている。

【0335】リソース管理部3201は、各ユーザ端末3208に対してデータを送出する際に、必要となるリソースを管理する。データ記憶部3202は、ランダムアクセスが可能な複数の記憶装置（例えば、ハードディスクドライブ、オートチェンジャー付き光ディスクドライブ）3205と、各記憶装置3205に接続されてデータの読み出しを制御する複数のMSFS3206とを含む。交換部3203は、入出力部3204とデータ記憶部3202との間でデータの交換を行う。なお、本実施形態では、一例として、交換部3203にATMスイッチを使用している。入出力部3204は、複数のI/Oユニット3207を含み、サーバ装置と各ユーザ端末3208との間のデータの入出力を制御する。各ユーザ端末3208は、いずれかのI/Oユニット3207に接続されている。

【0336】上記のような構成において、各I/Oユニット3207と各MSFS3206との間には、交換部3203によりメッシュ状に接続が設定される。なお、各接続の帯域は、リソース管理部3201が管理している。

【0337】図33は、図32におけるI/Oユニット3207のより詳細な構成を示すブロック図である。図33において、I/Oユニット3207は、優先入出力制御部3301と、非優先入出力制御部3302とを備え

ている。非優先入出力制御部3302は、コマンドキュー3321と、コマンドキュー制御部3322と、コマンド送出部3323と、データバッファ部3324と、バッファ管理部3325と、データ送出部3326と、コマンド受付部3327とを含む。

【0338】優先入出力制御部3301は、例えば動画データのように連続性を必要とするデータ（以下、優先データと称する）の入出力を制御する。これに対し、非優先入出力制御部3302は、例えば静止画データのように連続性を必要としないデータ（以下、非優先データと称する）の入出力を制御する。

【0339】非優先入出力制御部3302において、コマンドキュー3321は、ユーザ端末3208からの非優先データの読み出しを要求するコマンドを、ユーザ端末毎に蓄える。コマンドキュー制御部3322は、コマンドキュー3321からのコマンドの取り出しを制御する。コマンド送出部3323は、コマンドキュー制御部3322が取り出したコマンドを、対応する宛先のMSFS3206に送出する。データバッファ部3324は、MSFS3206が記憶装置3205から読み出したデータを蓄える。バッファ管理部3325は、データバッファ部3324に格納されている（または、格納が予約されている）データを、読み出しを要求したユーザ端末毎にカウントする。データ送出部3326は、データバッファ部3324に蓄積されたデータを、対応する宛先のユーザ端末3208に送出し、送出が完了したデータバッファ部3324の領域を空きの状態に設定する。コマンド受付部3327は、ユーザ端末3208から到着したコマンドを、コマンドキュー3321に登録するか否かを管理している。

【0340】以上のように構成されたサーバ装置について、以下にその動作を説明する。まず、システムの起動時に行われるリソースの初期設定動作について説明する。リソース管理部3201は、システムの起動時において、データ伝送システムにおける各部のリソース、すなわち、各MSFS3206のデータ読み出し帯域、各I/Oユニット3207における外部との間の通信帯域、交換部3203における各ATMスイッチの通信帯域等を、（1）優先データを出力するのに使用するリソース（以下、優先リソースと称す）と、（2）非優先データを出力するのに使用するリソース（以下、非優先リソースと称す）とに分離して割り当てる。以後、リソース管理部3201は、分離されたそれぞれのリソースを、個別に管理する。

【0341】なお、MSFS3206のデータ読み出し帯域は、MSFS3206が要求する、読み出すデータの格納先がランダムで、かつ要求間隔がランダムな読み出し要求にตอบสนองしてデータの読み出しを行うときに、予め定められた遅延時間以内に、予め定められた廃棄率以下で読み出し可能なデータの最大速度として定義され

る。

【0342】次に、優先データの出力動作について説明する。各ユーザ端末3208は、サーバ装置に対して優先データの出力を要求する場合、当該優先データを送出するのに必要な帯域のリソースの確保を要求する。リソース確保の要求を受けた優先入出力制御部3301は、リソース管理部3201に対し、帯域量を指定してリソースの確保を要求する。リソース管理部3201では、要求された帯域量のデータを送信するために、ATMスイッチの通信帯域と、MSFS3206のデータ読み出し帯域と、I/Oユニット3207の通信帯域とを、優先リソースとして予め割り当てられている帯域量の中から確保する。

【0343】現在、優先リソースが使用中で、要求された帯域量に見合う分の余りがない場合には、リソースの確保が失敗し、サーバ装置は、優先データの出力を実行できない。一方、リソースの確保に成功すると、ユーザ端末3208が出力を希望するデータ用に、I/Oユニット3207の通信帯域と、MSFS3206のデータ読み出し帯域と、交換部3203のATMスイッチの通信帯域とが、それぞれの機器に設定される。

【0344】ユーザ端末3208は、リソース確保完了通知を優先入出力制御部3301から受けると、優先データの出力を要求するコマンドを発行する。コマンドを受信したI/Oユニット3207の優先入出力制御部3301は、出力したい優先データ用に確保されているリソースを用いてデータの出力を実行する。

【0345】図34は、図33におけるコマンド受付部3327の動作を示すフローチャートである。図35は、図33におけるコマンドキュー制御部3322の動作を示すフローチャートである。図36は、図33におけるコマンド送出部3323の動作を示すフローチャートである。以下、これら図34～図36を参照して、非優先データの出力動作について説明する。

【0346】図34に示すように、コマンド受付部3327は、ユーザ端末3208から非優先データの読み出しを要求するコマンドを受信すると（ステップS101）、当該コマンドを送信したユーザ端末3208に対応するコマンドキュー3321に空きがあるか否かを判断する（ステップS102）。対応するコマンドキュー3321に空きがない場合、コマンド受付部3327は、コマンドの受付を拒否し（ステップS103）、ステップS101の動作に戻る。一方、対応するコマンドキュー3321に空きがある場合、コマンド受付部3327は、受信したコマンドを対応するコマンドキュー3321の最後部に格納し（ステップS104）、その動作を終了する。なお、非優先入出力制御部3302は、前述した優先入出力制御部3301のように、ユーザ端末3208にデータを送出するためのリソースを確保する必要はない。

【0347】図35に示すように、コマンドキュー制御部3322は、バッファ管理部3325に格納されている管理情報を参照することにより、コマンドが蓄えられているユーザ端末3208毎のコマンドキュー3321の中から、データバッファ部3324に格納されているデータ量（実際に格納されているデータ量のみならず、格納することが予約されているデータ量も含む）が最も少ない（データが全く格納されていない場合も含む）ユーザ端末3208に対応するコマンドキューを選択する（ステップS201）。なお、データバッファ部3324に格納されているデータ量が同一であり、しかもそれぞれのデータ量が最小であるようなユーザ端末が複数存在する場合は、キュー長が最も長い（すなわち、最も多くのコマンドを格納している）コマンドキューが選択される。さらにこの場合であって、キュー長が同一のコマンドキューが複数存在する場合は、デフォルトで定められた順番に従ってコマンドキューがサイクリックに選択される。

【0348】次に、コマンドキュー制御部3322は、選択したコマンドキューに注目し、その先頭部に格納されているコマンドをCOM(1)とする（ステップS202）。次に、コマンドキュー制御部3322は、コマンドカウンタNcに初期値1を設定すると共に、送出コマンドカウンタNsに初期値0を設定する（ステップS203）。次に、コマンドキュー制御部3322は、コマンドCOM(Nc)の宛先のMSFS3206が、コマンドCOM(1)の宛先のMSFS3206と等しいか否かを判断する（ステップS204）。両コマンドの宛先となるMSFS3206が等しい場合、コマンドキュー制御部3322は、選択したコマンドキューからコマンドCOM(Nc)を取り出し、取り出したコマンドCOM(Nc)をコマンド送出部3323に出力する（ステップS205）。なお、コマンドカウンタNcの計数値の初期値は1であるため、最初は、コマンドCOM(Nc)=COM(1)となる。従って、最初は、選択したコマンドキューから先頭のコマンドCOM(1)が取り出され、コマンド送出部3323に出力される。

【0349】次に、コマンドキュー制御部3322は、送出コマンドカウンタNsをインクリメントする（ステップS206）。次に、コマンドキュー制御部3322は、送出コマンドカウンタNsの計数値が予め定められたしきい値Nよりも小さいか否かを判断する（ステップS207）。ここで、送出コマンドカウンタNsの計数値は、コマンドキュー制御部3322が、選択したコマンドキューの中から1つのMSFS3206に対して送出したコマンドの数に対応している。1つのMSFS3206に対してほぼ同時に多数のコマンドが送出されると、コマンドを受信したMSFS3206で輻輳が生じるため、本実施形態では、1つのコマンドキュー3321からは、同時にN個以上のコマンドを送出しないよう

にしている。すなわち、コマンドキュー制御部3322は、送出コマンドカウンタNsの計数値が予め定められたしきい値Nよりも小さい場合のみ、選択したコマンドキュー内の次のコマンドに対する送出処理に移行することができる。

【0350】上記ステップS207において、送出コマンドカウンタNsの計数値が予め定められたしきい値Nよりも小さい場合、コマンドキュー制御部3322は、選択したコマンドキュー内にCOM(Nc+1)で示されるコマンドが存在するか否か、すなわち選択したコマンドキュー内に未だ送出すべきコマンドが存在するか否かを判断する（ステップS208）。選択したコマンドキュー内にコマンドCOM(Nc+1)が存在する場合、すなわち選択したコマンドキュー内に未だ送出すべきコマンドが存在する場合、コマンドキュー制御部3322は、コマンドカウンタNcをインクリメントした後（ステップS209）、ステップS204の動作に戻る。従って、選択したコマンドキュー内において、次の位置に格納されているコマンドが、コマンド送出部3323への送出処理の対象となる。例えば、選択したコマンドキュー内において2番目の位置に格納されたコマンドCOM(2)について考えてみると、このコマンドCOM(2)が先頭位置のコマンドCOM(1)と等しい場合は、当該コマンドCOM(2)がステップS205においてコマンド送出部3323に出力される。一方、コマンドCOM(2)が先頭位置のコマンドCOM(1)と等しくない場合は、ステップS204の後、ステップS205～S207がスキップされて、当該コマンドCOM(2)がコマンド送出部3323に出力されない。

【0351】一方、選択したコマンドキュー内にコマンドCOM(Nc+1)が存在しない場合、コマンドキュー制御部3322は、コマンド送出処理が完了したか否か、すなわちコマンド送出部3323に出力したコマンドに対応するデータがMSFS3206からデータバッファ部3324に到着したか否かを判断し（ステップS210）、完了した場合はステップS201の動作に戻る。なお、コマンドキュー制御部3322は、コマンド送出部3323から受信完了通知（後述する図36のステップS308参照）を受け取ったときに、コマンド送出処理が完了したものと判断する。

【0352】なお、前述のステップS207において、送出コマンドカウンタNsの計数値が予め定められたしきい値N以上になった場合、コマンドキュー制御部3322は、ステップS210の動作に進み、次のコマンドに対する送出処理を実行しない。

【0353】図36に示すように、コマンド送出部3323は、コマンドキュー制御部3322から非優先データの読み出しを要求するコマンドを受け取ると、読み出し完了カウンタNrを初期化する（ステップS30

1)。すなわち、コマンド送出部3323は、読み出し完了カウンタNrの初期値として、コマンドキュー制御部3322から受け取った全コマンド数を設定する。次に、コマンド送出部3323は、MSFS3206に対して送出していないコマンドを内部に保有しているか否かを判断する(ステップS302)。コマンド送出部3323は、未送出のコマンドを保有している場合、保有している未送出のコマンドの中からいずれか1つのコマンド(通常は、次のコマンド)を選択して注目し、当該コマンドに対応するデータを記憶装置3205から読み出したときに、当該データを格納し得るだけの空き領域がデータバッファ部3324内に存在するか否かを判断する(ステップS303)。データバッファ部3324内に空き領域が存在するか否かは、バッファ管理部3325内の管理情報(具体的には、各カウンタの計数値)を参照することにより判断できる。

【0354】データバッファ部3324内に空き領域が存在する場合、コマンド送出部3323は、データを格納するための領域をデータバッファ部3324内に予約すると共に、バッファ管理部3325に予約したことを通知し、さらに宛先のMSFS3206に選択したコマンドを送出する(ステップS304)。仮に、予約できない場合には、予約できる状態になるまで、コマンド送出部3323は、非優先データの読み出しを要求するコマンドをMSFS3206に送出しない。この処理を行うことにより、MSFS3206からのデータ受信時に、データバッファ部3324内に格納領域が無くて、データが廃棄されることを防止することができる。

【0355】バッファ管理部3325は、ユーザ端末毎にカウンタを有しており、コマンド送出部3323から予約の通知を受けたとき、データバッファ部3324に格納領域を予約したデータの読み出しコマンドを送信したユーザ端末3208に対応するカウンタの計数値をインクリメントする。

【0356】次に、コマンド送出部3323は、データバッファ部3324がMSFS3206からデータを受信したか否かを判断する(ステップS305)。データバッファ部3324がMSFS3206からのデータを受信していない場合、コマンド送出部3323は、ステップS302の動作に戻り、次の未送出コマンドに対する送出処理を行う。一方、データバッファ部3324がMSFS3206からのデータを受信した場合、コマンド送出部3323は、読み出し完了カウンタNrをデクリメントする(ステップS306)。従って、読み出し完了カウンタNrの計数値は、今回の送出処理の対象となるコマンド(すなわち、コマンドキュー制御部3322から今回受け取った全コマンド)の内、MSFS3206からのデータ送出が完了していないコマンドの数を表すことになる。

【0357】その後、コマンド送出部3323は、読み

出し完了カウンタNrの計数値が0になったか否かを判断する(ステップS307)。読み出し完了カウンタNrの計数値が0になっていない場合、コマンド送出部3323は、ステップS302の動作に戻り、次の未送出コマンドに対する送出処理を行う。一方、読み出し完了カウンタNrの計数値が0になった場合、コマンド送出部3323は、受信完了通知をコマンドキュー制御部3322に送出し(ステップS308)、その動作を終了する。

【0358】上記の説明から明らかなように、コマンド送出部3323は、コマンドキュー制御部3322から送られたコマンドを全て送出した後、送出したコマンドに対応するデータをMSFS3206から全て受信完了するまでは、次のコマンドをコマンドキュー制御部3322から受け取らないことになる。

【0359】非優先データの読み出しを要求するコマンドを受信したMSFS3206は、当該コマンドを、内部に設けられた非優先コマンドキューに一旦蓄える。そして、当該MSFS3206は、非優先リソースとして確保されている読み出し帯域内で、非優先コマンドキューからコマンドを取り出し、記憶装置3205からのデータの読み出しを実行する。読み出されたデータは、読み出しを要求したI/Oユニット3207がMSFS3206との間に確保している非優先リソースとしての通信帯域を用いて送信される。このとき、当該I/Oユニット3207は、非優先データの受信に関して、同時に複数のMSFS3206との間で通信することがないため、MSFS3206は、I/Oユニット3207が持つ非優先リソースの通信帯域の全てを使用してデータを送信することが可能となる。

【0360】MSFS3206からのデータがI/Oユニット3207に到着すると、当該データは、データバッファ部3324内の既に予約されている領域に格納される。データ送出部3326は、非優先リソースとして確保されているサーバ装置外部との間の通信帯域内でデータバッファ部3324からデータを取り出し、データが格納されていた領域を空き領域とする。同時に、データ送出部3326は、データの取り出しをバッファ管理部3325に通知する。通知を受けたバッファ管理部3325は、データバッファ部3324から取り出されたデータの読み出しコマンドを送信したユーザ端末3208に対応するカウンタをデクリメントする。そして、データ送出部3326は、対応するユーザ端末3208に、取り出したデータを送出する。

【0361】以上説明したように、第8の実施形態によれば、システムの各部のリソースを、優先データを出力するのに使用するリソース(優先リソース)と、非優先データを出力するのに使用するリソース(非優先リソース)とに予め分離して管理するようにしているので、非優先データの読み出しに使用するリソースを、優先リソ

ースとは独立して確保することができる。その結果、非優先データの読み出し量が増加しても、優先データの出力に影響を与えることがない。

【0362】ところで、各ユーザ端末3208がサーバ装置からの非優先データを受信できる能力には限界がある。それにもかかわらず、あるユーザ端末3208が、自己の受信能力を超える量の非優先データの読み出しを要求した場合、そのユーザ端末3208からの要求に回答して記憶装置3205から読み出されたデータは、データバッファ部3324の内部において大量に蓄積されることになる。その結果、サーバ装置では、その他のユーザ端末3208からの読み出し要求に対する処理が遅くなる。以下、このことを具体例を挙げてより詳細に説明する。

【0363】例えば、サーバ装置に10台のユーザ端末3208が接続されており、各ユーザ端末3208の受信能力がそれぞれ10Mbpsであり、サーバ装置の外部との間の通信帯域が全体として100Mbpsであった場合を想定する。この場合、ある1台のユーザ端末3208が、サーバ装置に対して100Mbpsを越える量の非優先データの読み出し要求を出したとすると、データバッファ部3324内は、読み出し要求を出した1つのユーザ端末3208に対する蓄積データで満杯になる。しかも、このときデータバッファ部3324内に蓄積されたデータは、10Mbpsの速度でしか減少していかない。このような状況下では、他のユーザ端末3208から非優先データの読み出し要求があったとしても、データバッファ部3324内に読み出しデータを格納し得る空き領域ができるまでは、当該他のユーザ端末3208からの要求に対する送信処理を実行することができない。これでは、即時性を要求されるような読み出し要求コマンドに迅速に対応することができない。しかも、この場合、サーバ装置全体としては100Mbpsの通信能力を有しているにもかかわらず、実際には10Mbpsの通信能力しか発揮していないことになり、限られた通信帯域を有効に活用することができない。

【0364】そこで、上記第8の実施形態では、バッファ管理部3325に格納されている管理情報を参照することにより、コマンドが蓄えられているユーザ端末3208毎のコマンドキュー3321の中から、データバッファ部3324に格納されているデータ量が最も少ないユーザ端末3208に対応するコマンドキューを選択し、当該選択したコマンドキューに蓄えられているコマンドの送出処理を優先的に実行するようにしているの、複数のユーザ端末3208が共通に使用している非優先リソースを、1つのユーザ端末3208が占有して使用することが少なくなり、サーバ装置が各ユーザ端末3208に対して有する通信能力を最大限に活用することができる。

【0365】なお、コマンドキュー制御部3322は、

データバッファ部3324に格納されている最も少ないデータ量が所定のしきい値を越えている場合、選択したコマンドキュー3321からのコマンド取り出しを中止するようにしてもよい。これによって、データバッファ部3324が特定のユーザ端末3208のデータで埋め尽くされ、他のユーザ端末3208の非優先のデータの読み出し要求が処理できなくなることをより一層効果的に防ぐことができる。

【0366】また、上記第8の実施形態では、コマンド送出部3323は、データバッファ部3324内にデータを格納するための領域を予約した後に、非優先データの読み出しを要求するコマンドをMSFS3206に送出するようにしている。すなわち、コマンド送出部3323は、格納領域を予約できない場合には、予約できるまで、非優先データの読み出しを要求するコマンドをMSFS3206に送出しない。このようなコマンド送出部3323を設けたことにより、MSFS3206からのデータ受信時に、データバッファ部3324内に格納領域がなくてデータが廃棄されることを防止できる。

【0367】さらに、上記第8の実施形態では、コマンド送出部3323は、コマンドキュー制御部3322から送られたコマンドを全て送出後、送出したコマンドに対応するデータをMSFS3206から全て受信完了するまでは、次のコマンドをコマンドキュー制御部3322から受け取らない。これによって、I/Oユニット3207は、非優先データの受信に関して、同時に複数のMSFS3206と通信することがないため、非優先データ同士のセルの衝突を考慮する必要がない。また、MSFS3206は、I/Oユニット3207が持つ非優先リソースの通信帯域の全てを使用して高速にデータを送信することが可能となる。

【0368】（第9の実施形態）図37は、本発明の第9の実施形態に係るサーバ装置を用いたデータ伝送システムの構成を示すブロック図である。図37において、本データ伝送システムは、リソース管理部3701と、データ記憶部3702と、交換部3703と、入出力部3704と、複数のユーザ端末3708とを備えている。なお、本実施形態のサーバ装置は、リソース管理部3701と、データ記憶部3702と、交換部3703と、入出力部3704とによって構成される。そして、当該サーバ装置と各端末3708は、有線または無線の伝送路を介して接続されている。

【0369】リソース管理部3701は、各ユーザ端末3708に対してデータを送出する際に、必要となるリソースを管理する。データ記憶部3702は、ランダムアクセスが可能な複数の記憶装置（例えば、オートチェンジャー付き光ディスクドライバ）3705と、各記憶装置3705に接続されてデータの読み出しを制御する複数のMSFS3706とを含む。交換部3703は、入出力部3704とデータ記憶部3702との間でデー

タの交換を行う。なお、本実施形態では、一例として、交換部3703にATMスイッチを使用している。入出力部3704は、複数のI/Oユニット3707を含み、サーバ装置と各ユーザ端末3708との間のデータの入出力を制御する。各ユーザ端末3708は、いずれかのI/Oユニット3707に接続されている。

【0370】上記のような構成において、各I/Oユニット3707と各MSFS3706との間には、交換部3703によりメッシュ状にコネクションが設定される。なお、各コネクションの帯域は、リソース管理部3701が管理している。

【0371】上記第9の実施形態は、以下の点で前述した第8の実施形態と相違する。まず、リソース管理部3701は、システム全体のリソースを、優先リソースと非優先リソースとに分離して管理するが、必要となる優先リソース量の増減に応じて、優先リソースとして割り当てる量と、非優先リソースとして割り当てる量とを変更する。

【0372】次に、交換部3703におけるATMスイッチの非優先リソースとしての通信帯域は、ABRモードによって確保されている。ここで、ABRとは、ATMでの通信の標準化を推進する業界団体である“The ATM Forum”で審議されているサービスカテゴリである。ABRを用いた通信では、輻輳が生じた場合に送信レートがATMレイヤで自動的に調整され、それによって輻輳が解消される。また、帯域に空きがある場合には、自動的に送信レートが上げられ、通信帯域の有効利用が図られる。ABRについては、ATM Forum Traffic Management Specification Version 4.0, October 1995, ATM Forum/95-0013R8等に記述されている。

【0373】図38は、図37におけるI/Oユニット3707のより詳細な構成を示すブロック図である。図38において、I/Oユニット3707は、優先入出力制御部3801と、非優先入出力制御部3802とを備えている。非優先入出力制御部3802は、コマンドキュー3841と、コマンドキュー制御部3842と、コマンド送出部3843と、データバッファ部3844と、データ送出部3846と、コマンド受付部3847とを含む。コマンド送出部3843は、各MSFS3706毎に設けられた、複数のカウンタ3848および複数の送出キュー3849を含む。

【0374】優先入出力制御部3801は、動画データ等の優先データの入出力を制御する。これに対し、非優先入出力制御部3802は、静止画データ等の非優先データの入出力を制御する。

【0375】非優先入出力制御部3802において、コマンドキュー3841は、ユーザ端末3708からの非優先データの読み出しを要求するコマンドを一括的に蓄

える。コマンドキュー制御部3842は、コマンドキュー3841からのコマンドの取り出しを制御する。コマンド送出部3843は、コマンドキュー制御部3842が取り出したコマンドを、対応する宛先のMSFS3706に送出する。データバッファ部3844は、MSFS3706が記憶装置3705から読み出したデータを蓄える。データ送出部3846は、データバッファ部3844に蓄積されたデータを、対応する宛先のユーザ端末3708に送出する。コマンド受付部3847は、ユーザ端末3708から到着したコマンドを、コマンドキュー3841に登録するか否かを管理している。

【0376】コマンド送出部3843において、各カウンタ3848は、送出したデータ読み出しのコマンドの内、未だMSFS3706からデータが到着していないコマンドの数を、各MSFS3706毎にカウントする。各送出キュー3849は、各MSFS3706毎に、送出するコマンドを蓄える。

【0377】図39は、図37におけるリソース管理部3701の動作を示すフローチャートである。図40は、図38におけるコマンド受付部3847の動作を示すフローチャートである。図41は、図38におけるコマンドキュー制御部3842の動作を示すフローチャートである。図42は、図38におけるコマンド送出部3843の動作を示すフローチャートである。以下、これら図39～図42を参照して、第9の実施形態のサーバ装置の動作について説明する。

【0378】まず、システムの起動時に行われるリソースの初期設定動作について説明する。リソース管理部3701は、システムの起動時において、データ伝送システム全体のリソース（各MSFS3706のデータ読み出し帯域、各I/Oユニット3707における外部との間の通信帯域、交換部3703における各ATMスイッチの通信帯域等を含む）の内、非優先リソースとして最低限確保すべき帯域量を決定する。本実施形態では、優先リソースは、非優先リソースよりも常に優先的に確保される。従って、非優先リソースとして最低限確保すべき帯域量を決めておかないと、優先データの読み出し要求が増えた場合、非優先リソースとして確保する帯域が無くなり、非優先データを送出することができなくなる。また、システムの起動時において、リソース管理部3701は、ABRモードによって、MSFS3706とI/Oユニット3707との間の、非優先リソースとしてのATM通信路を設定する。

【0379】次に、優先データの出力動作について説明する。各ユーザ端末3708は、サーバ装置に対して優先データの出力を要求する場合、当該優先データを送出するのに必要な帯域のリソースの確保を要求する。リソース確保の要求を受けた優先入出力制御部3801は、リソース管理部3701に対し、帯域量を指定してリソースの確保を要求する。リソース管理部3701では、

要求された帯域量のデータを送信するために、交換部3703におけるATMスイッチの通信帯域と、MSFS3206のデータ読み出し帯域と、I/Oユニット3707の通信帯域とを、優先リソースとして予め割り当てられている帯域量の中から確保する。

【0380】現在、優先リソースとして確保されている帯域の全てまたは大部分が使用中であり、要求された優先データを読み出して送出するのに必要な帯域量の余りがない場合、図39に示すように、リソース管理部3701は、優先リソースとして確保している帯域量の変更要求が発生したものと判断する（ステップS401）。その後、リソース管理部3701は、もし優先リソースの帯域量を増加するように変更した場合、非優先リソースに対し、前述の最低帯域量（システムの起動時に設定されている）を確保することが可能か否かを判断する（ステップS402）。もし、優先リソースの帯域量を増加すると、非優先リソースに対して最低帯域量を確保することができない場合、リソース管理部3701は、優先リソースの増加を拒否し（ステップS403）、ステップS401の動作に戻る。この場合、優先リソースの確保が失敗し、サーバ装置は、優先データの出力を実行できない。

【0381】一方、優先リソースの帯域量を増加しても非優先リソースに対して最低帯域量を確保することが可能な場合、リソース管理部3701は、変更後の優先リソースの残りとして割り当てられる非優先リソースの帯域量をMSFS3706、I/Oユニット3707に設定すると共に、カウンタ3848のしきい値Lをコマンド送出部3843に通知する（ステップS404）。なお、リソース管理部3701は、優先入出力制御部3801が優先データの出力を完了すると、対応する帯域量を確保している優先リソースから減少させる。

【0382】リソース管理部3701のその他の動作は、図32に示すリソース管理部3201と同様であり、その説明を省略する。

【0383】次に、非優先データの出力動作について説明する。図40に示すように、コマンド受付部3847は、ユーザ端末3708から非優先データの読み出しを要求するコマンドを受信すると（ステップS501）、コマンドキュー3841に空きがあるか否かを判断する（ステップS502）。コマンドキュー3841に空きがない場合、コマンド受付部3847は、コマンドを送信したユーザ端末3708に対してコマンドの受付を拒否し（ステップS503）、ステップS501の動作に戻る。一方、コマンドキュー3841に空きがある場合、コマンド受付部3847は、受信したコマンドをコマンドキュー3841の最後部に格納し（ステップS504）、その動作を終了する。

【0384】図41に示すように、コマンドキュー制御部3842は、コマンドキュー3841にコマンドが格

納されているか否かを判断し（ステップS601）、格納されている場合は、MSFS毎に設けられた送出キュー3849の内、1つでも満杯状態（すなわち、最後尾までコマンドを格納した状態）になったキューがあるか否かを判断する（ステップS602）。もし、1つでも満杯状態になったキューがある場合、コマンドキュー制御部3842は、コマンドが送出され、送出キュー3849に空きができるまで、コマンドキュー3841からのコマンドの取り出しを停止する。一方、満杯状態の送出キュー3849が1つもない場合、コマンドキュー制御部3842は、コマンドキュー3841の先頭からコマンドを1つ取り出し、当該コマンドを送出先のMSFS3706に対応する送出キュー3849に格納する（ステップS603）。

【0385】次に、コマンドキュー制御部3842は、所定のデータ量に相当するコマンドをコマンド送出部3843に送出したか否かを判断する（ステップS604）。ここで、所定のデータ量とは、その時点でI/Oユニット3707に割り当てられている非優先リソースであって、MSFS3706との間の通信帯域と、ユーザ端末3708との通信帯域との内、小さい方の通信帯域を用いて、予め定められた時間Tの間に送信可能なデータ量と定義される。所定のデータ量に相当するコマンドをコマンド送出部3843に送出していない場合、コマンドキュー制御部3842は、ステップS601の動作に戻り、コマンドキュー3841から次のコマンドを取り出してコマンド送出部3843に送出する。一方、所定のデータ量に相当するコマンドをコマンド送出部3843に送出している場合、コマンドキュー制御部3842は、時間Tの間、コマンドキュー3841からコマンド送出部3843へのコマンドの送出を停止する（ステップS605）。

【0386】例えば、I/Oユニット3707がサーバ装置外部にデータを出力するための非優先リソースとしての通信帯域がM（KB/s）であり、I/Oユニット3707がMSFS3706から読み出しデータを受信するための非優先リソースとしての通信帯域がN（KB/s）であり、1つの読み出し要求コマンドで読み出されるデータ量がS（KB）であり、 $M > N$ である場合、コマンドキュー制御部3842は、 S/N （sec）の周期でコマンドキュー3841からコマンドを取り出し、コマンド送出部3843に送ることになる。

【0387】図42に示すように、コマンド送出部3843は、コマンドの送出先MSFSを表す値iを初期化、すなわち1に設定する（ステップS701）。次に、コマンド送出部3843は、i番目のカウンタ3848のカウント値 $Co(i)$ が、リソース管理部3701によって設定されたしきい値L（前述のステップS404で設定されたしきい値）よりも小さいか否かを判断する（ステップS702）。カウント値 $Co(i)$ がし

きい値 L よりも小さい場合（すなわち、 $Co(i) < L$ の場合）、コマンド送出部3843は、 i 番目のカウンタ3848のカウント値 $Co(i)$ をインクリメントし（ステップS703）、 i 番目の送出キュー3849からコマンドを1つ取り出して対応する宛先のMSFS3706に送出する（ステップS704）。次に、コマンド送出部3843は、値 i をインクリメントする（ステップS705）。

【0388】その後、コマンド送出部3843は、データバッファ部3844がMSFS3706からの読み出しデータを受信したか否かを判断し（ステップS706）、受信した場合は、当該受信データを送信した先のMSFS3706に対応するカウンタ3848のカウント値をデクリメントする（ステップS707）。次に、コマンド送出部3843は、値 i がMSFS3706の総個数よりも大きいのかを判断する（ステップS708）。値 i がMSFS3706の総個数以下の場合、コマンド送出部3843は、ステップS702の動作に戻り、次の送出キュー3849に対してコマンドの送出処理を実行する。

【0389】ここで、前述のステップS702において、 i 番目のカウンタ3848のカウント値 $Co(i)$ がしきい値 L 以上の場合（すなわち、 $Co(i) \geq L$ の場合）、コマンド送出部3843は、対応する i 番目の送出キュー3849に対するコマンドの送出処理を行わない。これは、1つのMSFS3706に対して読み出し処理が集中し、当該MSFS3706内でコマンドが滞留するのを避けるためである。

【0390】ステップS708において、値 i がMSFS3706の総個数よりも大きい場合、コマンド送出部3843は、ステップS701の動作に戻る。従って、値 i が初期値1に設定され、上記一連の動作が繰り返される。このように、コマンド送出部3843は、複数の送出キュー3849を1番目から順番にサーチし、注目している送出キュー3849がコマンドを格納している場合は、コマンドを1つだけ取り出して対応する宛先のMSFS3706に送出するようにしている。これによって、各MSFS3706は、均等に読み出し処理を行うことになり、いずれか1つのMSFS3706に負荷が集中するのを回避できる。さらに、偶然にユーザ端末3708からの読み出し要求が1つのMSFS3706に偏った場合であっても、当該MSFS3706に対する未処理のコマンド（データの読み出しが実行されていないコマンド）がしきい値 L 以上になると、当該MSFS3706に対するコマンド送出処理を停止するようにしているので、このような不測の事態にも対応できる。

【0391】非優先データ読み出し要求を受信したMSFS3706は、当該要求を内部の非優先コマンドキューに一旦蓄る。そして、当該MSFS3706は、非優先リソースとして確保されている読み出し帯域内で、非

優先コマンドキューからコマンドを取り出し、データの読み出しを実行する。

【0392】読み出されたデータは、非優先リソースとして、読み出しを要求したI/Oユニット3707が、MSFS3706との間にABRモードで確保している通信帯域を用いて送信される。

【0393】MSFS3706からのデータがI/Oユニット3707に到着すると、データバッファ部3844に格納される。このとき、送信元MSFS3706に対応するカウンタ3848のカウント値がデクリメントされる（前述のステップS707）。

【0394】データ送出部3846は、非優先リソースとして確保されているサーバ装置外部との通信帯域に従い、データバッファ部3844からデータを取り出し、対応するユーザ端末3708に取り出したデータを送出する。

【0395】以上のように、上記第9の実施形態によれば、優先リソースとして使用されていないリソースを全て非優先リソースとして管理し、優先リソースを新たに増やす必要が生じた場合には、非優先リソースが所定のリソース量（最低帯域量）以下にならない条件下で、非優先リソースから優先リソースを割り当てるようにしているので、非優先データの読み出しが増加しても、優先データの入出力が影響を受けることがなく、しかも、使用されていないリソースが全て非優先リソースとして使用されるため、非優先データの読み出しが効率的に実行される。

【0396】また、上記第9の実施形態では、各MSFS3706が非同期で動作するため、各MSFS3706からI/Oユニット3707に送信されるデータの衝突が発生し、交換部3703におけるATMスイッチに輻輳が生じる恐れがある。しかしながら、第9の実施形態では、非優先リソースとしてのATMスイッチの通信帯域をABRモードを用いて確保し、データを読み出したMSFS3706は、ABRモードで確保している通信帯域を用いてデータを送信するようにしているので、たとえ輻輳が生じた場合でも、即座にMSFS3706からの送信レートが下がり、輻輳が解消される。従って、I/Oユニット3707は、常に正常にデータを受信できる。ただし、優先リソースとしての通信帯域の解放等により、通信帯域に空き帯域が生じた場合には、直ちに送信レートが上がり、帯域の有効利用が行われる。

【0397】また、上記第9の実施形態では、コマンドキュー制御部3842は、I/Oユニット3707がサーバ装置外部にデータを出力するための非優先リソースとしての通信帯域と、I/Oユニット3707がMSFS3706から読み出しデータを受信するための非優先リソースとしての通信帯域との内、小さい方の帯域に、データの読み出しコマンドが要求するデータ量が等しくなる周期で、コマンドキュー3841の先頭から順にコ

マンドを取り出して、コマンド送出部3843に送るようにしている。これによって、MSFS3706から送られてくるデータによって使用される平均帯域が、ATMスイッチがMSFS3706と通信するための非優先リソースとしての通信帯域と等しくなるか、またはそれよりも小さくなるため、MSFS3706のデータ送信によるATMスイッチの輻輳を抑制する効果が期待できる。さらに、MSFS3706から送られてくるデータによって使用される平均帯域が、サーバ装置外部にデータを出力するための非優先リソースとしての通信帯域とも等しくなるか、またはそれよりも小さくなるため、I/Oユニット3707内のデータバッファ部3844に蓄積されるデータ量が増大していくことを抑えることができる。

【0398】また、上記第9の実施形態では、コマンド送出部3843は、注目している送出キュー3849に対応するカウンタ3848のカウント値が、予め定められたしきい値未満である場合にのみ、当該注目している送出キュー3849内のコマンドをMSFS3706に送信するようにしているので、MSFS3706から同時に送信される可能性があるデータ数を制限していることになり、ATMスイッチの輻輳を抑制する効果が期待できる。

【0399】なお、第9の実施形態では、MSFS3706は、非優先リソースとして確保されている読み出し帯域内で、内部の非優先コマンドキューからコマンドを取り出し、データの読み出しを実行するとしたが、優先リソースに対するデータのI/O処理が行われていないときに、非優先コマンドキューからコマンドを取り出し、非優先データの読み出しを実行するとしてもよい。この場合、非優先リソースの帯域量が変化した場合の各MSFS3706に対する非優先リソースとしてのデータの読み出し帯域の通知は、必要がなくなる。

【0400】また、上記第9の実施形態では、カウンタ3848および送出キュー3849は、MSFS3706毎に設けるとしたが、これらを各I/Oユニット3707に1つずつ設けるようにし、送出先のMSFSを意識しない制御としても、第9の実施形態と同様にMSFS3706から同時に送信される可能性があるデータ数を制限していることになり、ATMスイッチの輻輳を抑制する効果が期待できる。ただし、ユーザ端末3708からのデータの読み出しが特定のMSFS3706にある程度集中した場合等には、読み出しの効率が悪くなることが考えられる。これは、要求が集中していないMSFS3706に対しては、要求を送信できるとすぐに応答が得られるはずであるが、要求が集中しているMSFSからのデータを受信しないと、カウンタ3848のカウント値のしきい値制御により、要求をMSFS3706に送信できなくなるためである。

【0401】一般的に、ファイルシステムの利用者は、

取得するファイルデータの範囲を指定して、ファイルシステム内のサーバに対してデータ送出要求を送る。サーバは、データ送出要求で指定された範囲のファイルデータをすべて利用者に送出する。従来のファイルシステムでは、利用者は、サーバの実行中のデータ送出処理を中断する操作は行えたが、実行中のデータ送出処理を一時停止し、一時停止したデータ送出処理を再開する操作を行うことはできなかった。

【0402】ところで、デジタル化された動画データ扱うファイルシステムでは、任意のデータ伝送速度によるデータの連続送出処理が不可欠となっている。動画データ扱うファイルシステムに関して、ISO/IEC標準13818-9のドラフトでは、1つの動画ファイルに関するデータの連続送出の一時停止および再開操作を定義している。このドラフトによれば、「DSM Stream Pause」によりデータ送出の一時停止を行い、「DSM Stream Resume」によりデータ送出再開を実行するものと定義している。

【0403】しかしながら、動画データ扱う従来のファイルシステムにおいて、ある伝送速度によるデータの連続送出処理を一時停止するポーズ、および一時停止したデータ送出処理を再開するポーズ解除に関して、ポーズ解除時に任意に指定したデータ伝送速度でデータ送出を再開させるポーズ制御方式は確立されていなかった。

【0404】また、従来のファイルシステムは、ポーズ解除でデータの送出を再開する場合、ポーズ解除後のデータ取得に必要な時間を考慮していなかった。そのため、ポーズ処理前後でデータの伝送速度が変化すると、連続した送出に必要なデータの不足が生じたり、取得したデータのバッファからの溢れが生じたり、データ送出再開までにファイルシステムが次のデータを取得する時間が必要となる等の問題があった。

【0405】そこで、データの連続送出処理の一時停止および再開を行うファイルシステムにおいて、ポーズ解除後の伝送速度を任意に指定した場合でも、データ送出時のデータ不足を生ずることがなく、かつ再開までの時間を短縮することのできる実施形態について以下に説明する。

【0406】（第10の実施形態）図43は、本発明の第10の実施形態に係るファイルシステムの構成を示すブロック図である。図43において、本ファイルシステムは、サーバ装置4301と、データ記憶装置4302とを備えている。

【0407】データ記憶装置4302は、サーバ装置4301によって管理され、利用者4303によって利用されるデータを格納している。利用者4303は、指定した範囲のデータの送出を要求するデータ送出要求と、データ送出の一時停止を要求するポーズ要求と、一

時停止したデータ送出の再開を要求するポーズ解除要求とを、サーバ装置4301に対して発行する。サーバ装置4301は、利用者4303からのデータ送出要求を受けて、データ記憶装置4302からデータを取得し、当該データを利用者4303に送出するものであり、データ送出部4310と、データ取得部4311と、バッファ4312と、バッファ管理部4313と、ポーズ制御部4314とを含む。

【0408】データ取得部4311は、指定した位置から順番に、かつ指定された速度で連続的に、データ記憶装置4302からデータを取得する。バッファ4312は、データ記憶装置4302から取得したデータを格納する。バッファ管理部4313は、バッファ4312に格納されたデータの操作とデータ量の管理とを行う。データ送出部4310は、バッファ4312に格納されているデータを、指定された順番で、かつ指定された任意の伝送速度で連続的に、データを利用者4303に送出する。ポーズ制御部4314は、利用者4303からのポーズ要求およびポーズ解除要求を受けて、データ送出部4310とデータ取得部4311とバッファ管理部4313とを制御する。また、サーバ装置4301およびデータ記憶装置4302で扱うデータは、すべて固定長のブロックを単位として操作され、各ブロックにはデータ中の位置を示すブロック番号が連続的に付与されるものとする。

【0409】図44は、サーバ装置4301のデータ送出動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。まず、この図44を参照して、サーバ装置4301が利用者4303からデータ送出要求を受け付けたときの動作を説明する。

【0410】図44において、参照番号4441および4442は、時間軸を示しており、当該時間軸は、垂直方向の対応する位置が、互いに同一時刻を表している。参照番号4431は、サーバ装置4301が利用者4303から受け付けたデータ送出要求と、その受付時刻とを示している。参照番号4401A~4403Aは、データ取得部4311がデータ記憶装置4302に対して発行するデータ取得要求と、その発行時刻とを示している。参照番号4411A~4413Aは、データ取得要求4401A~4403Aによって取得されたデータと、データ送出部4310が利用者4303へ当該データを送出する時刻とを示している。参照番号4421は、データ取得要求4401Aを発行してから対応するデータ4411Aがバッファ4312に格納されるまでの最大時間を示している。参照番号4451~4454は、データ送出部4310が利用者4303にデータを送出する時間間隔を示している。

【0411】利用者4303は、データ送出要求4431を発行し、データの範囲とデータを送出する際の伝送

速度とを指定する。データの範囲は、開始ブロック番号と終了ブロック番号とで指定される。データの伝送速度は、送出するブロック間の平均時間間隔であるデータ送出間隔（以降Tsdと記述する）で指定される。

【0412】データ取得部4311は、データ送出要求4431を受け付けると、当該データ送出要求4431で指定された開始ブロック番号に対応する1つのブロックを取得するためのデータ取得要求4401Aを発行する。データ取得部4311は、データ取得要求4401Aの発行時刻からデータ送出間隔Tsd後に、次の1つのブロックを取得するためのデータ取得要求4402Aを、データ記憶装置4302に発行する。同様の手順により、データ取得部4311は、データ送出要求4431で指定された終了ブロック番号に到達するまで、データ送出要求を繰り返し発行する。

【0413】データ取得要求4401Aによってデータ記憶装置4302から取得されたブロックは、バッファ4312に格納される。データ取得時間Dは、データ取得部4311がデータ取得要求を発行してから対応する1つのブロックがバッファ4312に格納されるまでの最大時間であり、既知の値である。すなわち、データ取得時間Dの間に、必ず指定した1ブロックを取得できることが保証される。

【0414】データ取得要求4401Aによって取得され、バッファ4312に格納されたブロックは、データ取得要求4401Aの発行時刻からデータ取得時間D経過後のデータ送出時刻4411Aに、データ送出部4310により利用者4303に対して送出される。同様に、データ取得要求4402Aによって取得されたブロックは、データ取得要求4402Aの発行時刻からデータ取得時間D経過後の送出時刻4412Aに、データ送出部4310により利用者4303に対して送出される。同様の手順により、データ送出要求4431で指定された終了ブロック番号に対応するデータが送出されるまで、データ送出が連続的に行われる。以上の動作により、利用者4303に指定された任意のデータ伝送速度による連続的なデータ送出が実現される。

【0415】図45は、サーバ装置4301のポーズ動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。次に、図45を参照して、サーバ装置4301内のポーズ制御部4314が、利用者4303からポーズ要求を受け取ったときの動作を説明する。

【0416】図45において、参照番号4432は、ポーズ制御部4314が利用者4303から受け付けたポーズ要求と、その受付時刻とを示している。参照番号4404A~4406Aは、データ取得部4311がデータ記憶装置4302に対して発行するデータ取得要求と、その発行時刻とを示している。参照番号4414Aおよび4415Aは、データ取得要求4404Aおよび

4045Aによって取得されたデータと、データ送出部4310が利用者4303に当該データを送出する時刻とを示している。参照番号4422は、データ取得要求4404Aを発行してから対応するデータ4414Aがバッファ4312に格納されるまでの最大時間Dを示している。参照番号4455～4457は、データ送出部4310が利用者4303にデータを送出する時間間隔を示している。

【0417】利用者4303からのデータ送出要求にตอบสนองして、データ取得部4311がデータ取得要求4404A～4406Aを発行しようとしているときであって、データ送出要求4404Aおよび4405Aが発行された後に、ポーズ制御部4314がポーズ要求4432を受け付けた場合、ポーズ制御部4314は、データ取得部4311に対して、直ちにデータ取得要求の発行を停止するように制御する。これにより、データ取得要求4406Aは、データ取得部4311から発行されなくなる。同様に、ポーズ制御部4314は、データ送出部4310に対して、直ちにデータ送出を停止するように制御する。データ取得要求4404Aおよび4405Aによって取得されたブロック4414Aおよび4415Aの送出時刻は、ポーズ要求受付時刻よりも後であるため、当該ブロック4414Aおよび4415Aは、バッファ4312に格納されたままとなり、データ送出部4310から利用者4303に送出されなくなる。以上の動作により、連続的に送出されていたデータの一時停止が実現される。

【0418】図46は、サーバ装置4301のポーズ解除動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。次に、図46を参照して、ポーズ制御部4314が利用者4303からポーズ解除要求を受け取ったときの動作を説明する。

【0419】図46において、参照番号4433は、ポーズ制御部4314が利用者4303から受け付けたポーズ解除要求と、その受付時刻とを示している。参照番号4406A～4408Aは、データ取得部4311がデータ記憶装置4302に対して発行するデータ取得要求と、その発行時刻とを示している。参照番号4414Aおよび4415Aは、ポーズ処理でバッファに格納されたままとなっている2つのブロックと、利用者4303への送出時刻とを示している。参照番号4416Aおよび4417Aは、ポーズ解除後のデータ取得要求4406Aおよび4407Aによって取得されたデータと、データ送出部4310が利用者4303に当該データを送出する時刻とを示している。参照番号4423は、データ取得要求4406Aを発行してから対応するデータ4416Aがバッファ4312に格納されるまでの最大時間Dを示している。参照番号4458～4462は、データ送出部4310が利用者4303に対してデータ

を送出する時間間隔を示している。

【0420】利用者4303は、ポーズ解除要求4433を発行し、ポーズ解除後のデータ伝送速度を指定する。データ伝送速度は、データ送出要求時と同様に、送出するブロック間の平均時間間隔である、データ送出間隔 T_{sd} で指定する。

【0421】ポーズ制御部4314は、利用者4303からポーズ解除要求4433を受け付けると、バッファ管理部4313を参照することにより、ポーズ処理以降にバッファ4312に格納されたままとなっているブロックの数Nを取得する。

【0422】ポーズ解除要求4433にตอบสนองして、ポーズ制御部4314がデータ取得部4311に対して、直ちにデータ取得を再開するよう制御する時刻を、データ取得再開時刻 t_{rd} と定義する。同様に、バッファ4312中に格納されているブロックの中で、次に送出するブロックの送出時刻、すなわち図46のデータ送出時刻4414Aを、データ送出再開時刻 t_{sd} と定義する。

【0423】ポーズ制御部4314は、データ取得再開時刻 t_{rd} と、バッファ4312中のブロック数Nと、ポーズ解除後のデータ送出間隔 T_{sd} と、データ取得時間Dとを、次式(14)に代入することにより、データ送出再開時刻 t_{sd} を求める。

$$t_{sd} = t_{rd} + D - N \cdot T_{sd} \quad \dots (14)$$

【0424】次に、ポーズ制御部4314は、上式(14)を用いて求めたデータ送出再開時刻 t_{sd} と、データ取得再開時刻 t_{rd} とを比較し、当該データ送出再開時刻 t_{sd} がデータ取得再開時刻 t_{rd} 以上の場合は、式(14)で求めた時刻を、データ送出再開時刻 t_{sd} と決定する。また、データ送出再開時刻 t_{sd} がデータ取得再開時刻 t_{rd} 未満の場合は、データ取得再開時刻 t_{rd} と等しい時刻を、データ送出再開時刻 t_{sd} と決定する。

【0425】次に、ポーズ制御部4314は、データ取得再開時刻 t_{rd} において、データ取得部4311に対し、バッファ4312に格納されている最後のブロック4415Aの次のブロック4416Aを取得するためのデータ取得要求4406Aから順番に、終了ブロックに到達するまで、データ取得要求を発行するよう制御する。同様に、ポーズ制御部4314は、決定したデータ送出再開時刻 t_{sd} において、データ送出部4310に対して、直ちにバッファ4312に格納されている先頭のブロック4414Aから順番に終了ブロックを送出するまで、データを送出するよう制御する。

【0426】以上のように、第10の実施形態によれば、ポーズ解除後のデータ伝送速度として任意の伝送速度が指定された場合、ポーズ制御部4314が、前述の式(14)を用いてデータ送出再開時刻を求めることにより、ポーズ解除後にデータが途切れることなく、指定された伝送速度でデータを送出することができる。

【0427】(第11の実施形態)以下、本発明の第1

1の実施形態に係るファイルシステムについて図面を参照しながら説明する。なお、第11の実施形態は、図43に示す第10の実施形態と各ブロックの機能が異なるが、全体としての構成は同様である。従って、以下には、図43に示す第10の実施形態の構成を援用しつつ、第11の実施形態について説明する。

【0428】第11の実施形態において、サーバ装置4301が利用者4303からデータ送出要求を受け付けたときの動作は、第10の実施形態において、図44を参照して説明した動作と同一である。同様に、ポーズ制御部4314が利用者4303からポーズ要求を受け付けたときの動作は、第10の実施形態において、図45を参照して説明した動作と同一である。従って、これらの動作の説明は、省略する。

【0429】図47は、サーバ装置4301のポーズ解除動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。以下、図47を参照して、ポーズ制御部4314が利用者4303からポーズ解除要求を受け付けたときの動作を説明する。

【0430】図47において、参照番号4741および4742は、時間軸を示しており、当該時間軸は、垂直方向の対応する位置が、互いに同一時刻を表している。参照番号4733は、ポーズ制御部4314が利用者4303から受け付けたポーズ解除要求と、その受付時刻とを示している。参照番号4706Bおよび4707Bは、データ取得部4311がデータ記憶装置4302に対して発行するデータ取得要求と、その発行時刻とを示している。参照番号4714Bおよび4715Bは、ポーズ処理でバッファ4312に格納されたままとなっている2つのブロックと、利用者4303への送出時刻とを示している。参照番号4716Bおよび4717Bは、データ取得要求4706Bおよび4707Bによって取得されたデータと、データ送出部4310が利用者4303に当該データを送出する時刻とを示している。参照番号4724は、データ取得要求4706Bを発行してから対応するデータ4716Bがバッファ4312に格納されるまでの最大時間Dを示している。4763～4766は、データ送出部4310が利用者4303にデータを送出する時間間隔を示している。

【0431】利用者4303は、ポーズ解除要求4733を発行し、ポーズ解除後のデータ伝送速度を指定する。データ伝送速度は、データ送出要求時と同様に、送出するブロック間の平均時間間隔であるデータ送出間隔 T_{sd} で指定する。

【0432】ポーズ制御部4314は、利用者4303からポーズ解除要求4733を受け付けると、バッファ管理部4313を参照して、ポーズ処理以降にバッファ4312に格納されたままとなっているブロックの数Nを取得する。

【0433】ポーズ解除要求4733に応答して、ポーズ制御部4314がデータ取得部4311に対して直ちにデータ取得再開するよう制御する時刻を、データ取得再開時刻 t_{rd} と定義する。同様に、バッファ4312中に格納されているブロックの中で、次に送出するブロックの送出時刻、すなわち図47のデータ送出時刻4714Bを、データ送出再開時刻 t_{sd} と定義する。

【0434】ポーズ制御部4314は、次式(15)に、データ送出再開時刻 t_{sd} と、バッファ4312中のブロック数Nと、ポーズ解除後のデータ送出間隔 T_{sd} と、データ取得時間Dとを代入することにより、データ取得再開時刻 t_{rd} を求める。

$$t_{rd} = t_{sd} + N \cdot T_{sd} - D \quad \dots (15)$$

【0435】ポーズ制御部4314は、上式(15)を用いて求めたデータ取得再開時刻 t_{rd} と、データ送出再開時刻 t_{sd} とを比較し、当該データ取得再開時刻 t_{rd} がデータ送出再開時刻 t_{sd} 以上の場合は、式(15)で求めた時刻を、データ取得再開時刻 t_{rd} と決定する。また、データ取得再開時刻 t_{rd} がデータ送出再開時刻 t_{sd} 未満の場合は、第10の実施形態を適用して、ポーズ制御部4314は、式(14)に、データ取得再開時刻 t_{rd} と、バッファ4312中のブロック数Nと、ポーズ解除後のデータ送出間隔 T_{sd} と、データ取得時間Dを代入することにより、データ送出再開時刻 t_{sd} を求める。

【0436】ポーズ制御部4314は、決定したデータ取得再開時刻 t_{rd} において、データ取得部4311に対し、バッファに格納されている最後のブロック4715Bの次のブロック4716Bを取得するためのデータ取得要求4706Bから順番に、終了ブロックに到達するまで、データ取得要求を発行するよう制御する。同様に、ポーズ制御部4314は、データ送出再開時刻 t_{sd} において、データ送出部4310に対し、直ちにバッファ4312に格納されている先頭のブロック4714Bから順番に、終了ブロックを送出するまで、データを送出するよう制御する。

【0437】以上のように、第11の実施形態によれば、ポーズ解除後のデータ伝送速度として任意の伝送速度が指定された場合、ポーズ制御部4314が、前述の式(15)を用いてデータ取得再開時刻を求めることにより、ポーズ解除後にデータが途切れることなく、指定された伝送速度でデータを送出することができる。

【0438】(第12の実施形態)以下、本発明の第12の実施形態に係るファイルシステムについて図面を参照しながら説明する。なお、第12の実施形態は、図43に示す第10の実施形態と各ブロックの機能が異なるが、全体としての構成は同様である。従って、以下には、図43に示す第10の実施形態の構成を援用しつつ、第12の実施形態について説明する。

【0439】第12の実施形態において、サーバ装置4301が利用者4303からデータ送出要求を受け付け

たときの動作は、第10の実施形態において、図44を参照して説明した動作と同一である。同様に、ポーズ制御部4314が利用者4303からポーズ要求を受け取ったときの動作は、第10の実施形態において、図45を参照して説明した動作と同一である。従って、これらの動作の説明は、省略する。

【0440】図48は、ポーズ解除動作におけるデータ送出再開時刻の算出方法を説明するためのタイムチャートである。図49は、サーバ装置4301のポーズ解除動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。以下、これら図48および図49を参照して、ポーズ制御部4314が利用者4303からポーズ解除要求を受け付けたときの動作を説明する。

【0441】図48において、参照番号4841および4842は、時間軸を示しており、当該時間軸は、垂直方向の対応する位置が、互いに同一時刻を表している。参照番号4833は、ポーズ制御部4314が利用者4303から受け付けたポーズ解除要求と、その受付時刻とを示している。参照番号4806Cは、データ取得部4311がデータ記憶装置4302に対して発行するデータ取得要求と、その発行予定時刻とを示している。参照番号4814Cおよび4815Cは、ポーズ処理でバッファ4312に格納されたままとなっている2つのブロックと、利用者4303への送出予定時刻とを示している。参照番号4816Cは、データ取得要求4806Cによって取得される予定のデータと、データ送出部4310が利用者4303に当該データを送出する予定時刻とを示している。参照番号4825は、データ取得要求4806Cを発行してから、対応するデータ4816Cがバッファ4312に格納されるまでの最大時間Dを示している。参照番号4867および4868は、データ送出部4310が利用者4303にデータを送出する時間間隔を示している。

【0442】図49において、参照番号4833は、ポーズ制御部4314が利用者4303から受け付けたポーズ解除要求と、その受付時刻とを示している。参照番号4805C'、4806Cおよび4807Cは、データ取得部4311がデータ記憶装置4302に対して発行するデータ取得要求の発行時刻を示している。参照番号4814C'は、ポーズ処理以降にバッファ4312に格納されたままとなっているブロックと、利用者4303への送出時刻とを示している。参照番号4815C'および4816Cは、データ取得要求4805C'および4806Cによって取得されたデータと、データ送出部4310が利用者4303に当該データを送出する時刻とを示している。参照番号4826は、データ取得要求4805C'を発行してから、対応するデータ4815C'がバッファ4312に格納されるまでの最大時間Dを示している。参照番号4869～4872は、

データ送出部4310が利用者4303にデータを送出する時間間隔を示している。

【0443】利用者4303は、ポーズ解除要求4833を発行し、ポーズ解除後のデータ伝送速度を指定する。データ伝送速度は、データ送出要求時と同様に、送出するブロック間の平均時間間隔である、データ送出間隔 T_{sd}' で指定する。

【0444】ポーズ制御部4314は、利用者4303からポーズ解除要求4833を受け付けると、バッファ管理部4313を参照して、ポーズ処理以降にバッファ4312に格納されたままとなっているブロックの数Nを取得する。

【0445】ポーズ解除要求4833に応答して、ポーズ制御部4314がデータ取得部4311に対して、直ちにデータ取得を再開するよう制御する時刻を、データ取得再開時刻 t_{rd} と定義する。同様に、バッファ4312中に格納されているブロックの中で、次に送出するブロックの送出時刻、すなわち図48のデータ送出時刻4814Cおよび図49のデータ送出時刻4814C'を、データ送出再開時刻 t_{sd} と定義する。

【0446】ポーズ制御部4314は、前述の式(14)を用いて、データ送出再開予定時刻 t_{sd} を求める。次に、ポーズ制御部4314は、式(14)を用いて求めたデータ送出再開予定時刻 t_{sd} と、データ取得再開時刻 t_{rd} とを比較し、当該データ送出再開予定時刻 t_{sd} がデータ取得再開時刻 t_{rd} 以上の場合は、式(14)で求めた時刻を、データ取得再開時刻 t_{sd} と決定し、前述した第10の実施形態で説明したポーズ解除処理と同様に、以下の処理を行う。すなわち、ポーズ制御部4314は、データ取得再開時刻 t_{rd} において、データ取得部4311に対し、直ちにバッファ4312に格納されている最後のブロックの次のブロックを取得するためのデータ取得要求から順番に、終了ブロックに到達するまで、データ取得要求を発行するよう制御する。同様に、ポーズ制御部4314は、決定したデータ送出再開時刻 t_{sd} において、データ送出部4310に対し、直ちにバッファ4312に格納されている先頭のブロックから順番に、終了ブロックを送出するまで、データを送出するよう制御する。

【0447】式(14)を用いて求めたデータ送出再開予定時刻 t_{sd} が、データ取得再開時刻 t_{rd} 未満、すなわち図48に示す場合は、以下の処理を行う。すなわち、ポーズ制御部4314は、図48のデータ取得時間4825において、ポーズ解除後のデータ伝送速度 T_{sd}' により送出することのできるブロック数L、すなわち限界データ量を、次式(16)を用いて求める。ただし、次式(16)において、[X]はXを越えない最大の整数とする。

$$L = [D / T_{sd}'] \quad \dots (16)$$

【0448】次に、ポーズ制御部4314は、バッファ

4312から削除するブロック数 N_d を、次式(17)を用いて求め、バッファ管理部4313に対し、バッファ4312に格納されているブロックから、次式(17)で求めた N_d 個のブロックを削除するよう制御する。

$$N_d = N - L \quad \dots (17)$$

【0449】バッファ管理部4313は、バッファ4312から N_d 個のブロックを、最後にバッファに格納された順番、すなわちブロック番号の大きい順番に削除する。これにより、バッファ4312中のブロック数は、 L と一致する。この結果、図48のブロック4815Cが、バッファ4312から削除される。

【0450】ポーズ制御部4314は、次式(18)に、データ取得再開時刻 t_{rd} と、バッファ4312中のブロック数 L と、ポーズ解除後のデータ送出間隔 $T_{sd'}$ と、データ取得時間 D とを代入することにより、新たにデータ送出再開時刻 $t_{sd'}$ を求める。

$$t_{sd'} = t_{rd} + D - L \cdot T_{sd'} \quad \dots (18)$$

【0451】次に、ポーズ制御部4314は、データ取得再開時刻 t_{rd} において、データ取得部4311に対し、バッファ4312に格納されている最後のブロック4814C'の次のブロック4815C'を取得するためのデータ取得要求4805C'から順番に、終了ブロックに到達するまで、データ取得要求を発行するよう制御する。同様に、ポーズ制御部4314は、決定したデータ送出再開時刻 $t_{sd'}$ において、データ送出部4310に対し、直ちにバッファ4312に格納されている先頭のブロック4814C'から順番に、終了ブロックを送出するまで、データを送出するよう制御する。

【0452】以上のように、第12の実施形態によれば、ポーズ解除後のデータ伝送速度として任意の伝送速度が指定された場合、限界データ量を越えるデータをバッファ4312から削除するようにしているので、サーバ装置4301は、限界データ量以上のバッファ容量を持つ必要がなく、ポーズ解除後にデータが途切れることなく、指定された伝送速度でデータを送出することができる。

【図面の簡単な説明】

【図1】本発明の第1の実施形態に係るファイルシステムの構成を示すブロック図である。

【図2】図1の各サーバユニットにおけるファイルアクセスに関連する部分の構成を示すブロック図である。

【図3】図1のファイルマネージャにおけるファイルアクセスに関連する部分の構成を示すブロック図である。

【図4】第1の実施形態において、リソースの予約および割り当てと、データ読み出し動作との関係を示した図である。

【図5】MS到着スロットの状態を管理するMS到着スロット管理テーブルの一例を示す図である。

【図6】書き込み用スロットの状態を管理する書き込み

用スロット管理テーブルの一例を示す図である。

【図7】ストリーム管理テーブルの一例を示す図である。

【図8】図3に示すリソース管理部のより詳細な構成を示すブロック図である。

【図9】本発明の第2の実施形態に係るファイルシステムの構成を示すブロック図である。

【図10】図9の各I/Oユニットにおけるファイルアクセス部の構成を示すブロック図である。

【図11】図9の各ブロック格納ユニットにおけるファイルアクセス部の構成を示すブロック図である。

【図12】本発明の第3の実施形態に係るファイルシステムの構成を示すブロック図である。

【図13】本発明の第4の実施形態に係るディスク管理装置の構成を示すブロック図である。

【図14】記憶装置の帯域に比例してデータを分配する比率を決定する、帯域優先レイアウトについて示す図である。

【図15】記憶装置の容量に比例してデータを分配する比率を決定する、容量優先レイアウトについて示す図である。

【図16】帯域優先レイアウトと容量優先レイアウトとを混在させてデータを分配する比率を決定する、帯域設定レイアウトについて示す図である。

【図17】図13に示すディスク管理装置を用いたファイルシステムの構成を示すブロック図である。

【図18】本発明の第6の実施形態に係るスケジューリング装置の構成を示すブロック図である。

【図19】図18のスケジューリング装置を用いた映像情報提供システムの構成を示すブロック図である。

【図20】図19の映像情報提供装置1920とクライアント(1~n)1921との間で、信号が送受信される様子を示すシーケンス図である。

【図21】図18の実行権発生周期テーブル1812の登録内容を示す図である。

【図22】図18の処理待ちキュー1814のキューイング内容を示す図である。

【図23】図18の実行権キュー1816のキューイング内容を示す図である。

【図24】図18の実行権発生部1813が実行権を発生させる動作を説明するためフローチャートである。

【図25】図18の実行指示部1815が読み出し要求の実行を指示する動作を説明するためのフローチャートである。

【図26】本発明の第7の実施形態に係るスケジューリング装置の構成を示すブロック図である。

【図27】図26のスケジューリング装置を用いた映像情報提供システムの構成を示すブロック図である。

【図28】図26の実行権発生周期テーブル2692の登録内容を示す図である。

【図29】図26の処理待ちキュー2694のキューイング内容を示す図である。

【図30】図26の実行権キュー2696のキューイング内容を示す図である。

【図31】他の実行権キューの動作を説明するための図である。

【図32】本発明の第8の実施形態に係るサーバ装置を用いたデータ伝送システムの構成を示すブロック図である。

【図33】図32におけるI/Oユニット3207のより詳細な構成を示すブロック図である。

【図34】図33におけるコマンド受付部3327の動作を示すフローチャートである。

【図35】図33におけるコマンドキュー制御部3322の動作を示すフローチャートである。

【図36】図33におけるコマンド送出部3323の動作を示すフローチャートである。

【図37】本発明の第9の実施形態に係るサーバ装置を用いたデータ伝送システムの構成を示すブロック図である。

【図38】図37におけるI/Oユニット3707のより詳細な構成を示すブロック図である。

【図39】図37におけるリソース管理部3701の動作を示すフローチャートである。

【図40】図38におけるコマンド受付部3847の動作を示すフローチャートである。

【図41】図38におけるコマンドキュー制御部3842の動作を示すフローチャートである。

【図42】図38におけるコマンド送出部3843の動作を示すフローチャートである。

【図43】本発明の第10の実施形態に係るファイルシステムの構成を示すブロック図である。

【図44】第10の実施形態において、サーバ装置のデータ送出動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。

【図45】第10の実施形態において、サーバ装置のポーズ動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。

【図46】第10の実施形態において、サーバ装置のポーズ解除動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。

【図47】第11の実施形態において、サーバ装置のポーズ解除動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。

【図48】第12の実施形態において、サーバ装置のポーズ解除動作における、データ送出再開時刻の算出方法を説明するためのタイムチャートである。

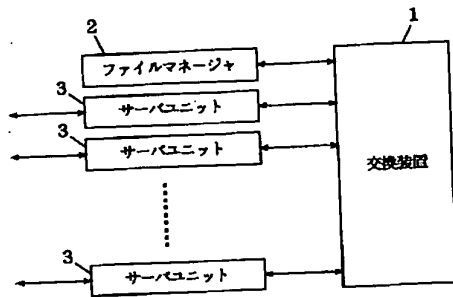
【図49】第12の実施形態において、サーバ装置のポーズ解除動作における、データ送出要求とデータ取得要求とデータ送出との関係を示すタイムチャートである。

【符号の説明】

- 1 交換装置
- 2 ファイルマネージャ
- 3 サーバユニット
- 101 外部メッセージインターフェース部
- 102 リソース割り当て部
- 103 スロット割り当て部
- 104 アドレス検索部
- 105 要求送出部
- 106 外部ネットワークI/O制御部
- 107 内部メッセージインターフェース部
- 108 記憶装置
- 109 記憶制御部
- 110 内部ネットワークI/O制御部
- 111 申告値管理部
- 201 メッセージインターフェース部
- 202 リソース管理部
- 401 申告パラメータ処理部
- 402 書き込み帯域算出部
- 403 受付判断部
- 404 リソース記憶部
- 901 ファイルマネージャ
- 902 I/Oユニット
- 905 ブロック格納ユニット
- 908 交換装置
- 1001 外部メッセージインターフェース部
- 1002 リソース割り当て部
- 1003 スロット割り当て部
- 1004 アドレス検索部
- 1005 要求送出部
- 1006 外部ネットワークI/O制御部
- 1007 内部メッセージインターフェース部
- 1008 申告値管理部
- 1101 内部メッセージインターフェース部
- 1102 内部ネットワークI/O制御部
- 1103 記憶制御部
- 1104 記憶装置
- 1201 メッセージインターフェース部
- 1202 リソース管理部
- 1203 読み出し、書き込み制御部
- 1204 記憶装置
- 1205 申告値管理部
- 1300 ディスク管理装置
- 1301 ディスク帯域管理部
- 1302 ディスク容量管理部
- 1303 分配比率決定部
- 1304 帯域計算部
- 1305 容量計算部
- 1306 全ディスク帯域制御部
- 1307 全ディスク容量制御部

1708a, 1708b 記憶装置
 1709 ファイル属性管理部
 1710 帯域管理部
 1711 ファイルオープン処理部
 1712 帯域獲得部
 1713 入出力制御部
 1414a~1414c 記憶装置
 1515a~1515c 記憶装置
 1616a~1616c 記憶装置
 1810, 2690 スケジューリング装置
 1811, 2691 登録指示部
 1812, 2692 実行権発生周期テーブル
 1813, 2693 実行権発生部
 1814, 2694 処理待ちキュー
 1815, 2695 実行指示部
 1816, 2696 実行権キュー
 1920, 2700 映像情報提供装置
 1921, 2701 クライアント
 1922, 2702 サーバ
 2703 伝送路
 3201, 3701...リソース管理部
 3202, 3702...データ記憶部
 3203, 3703...交換部

【図1】

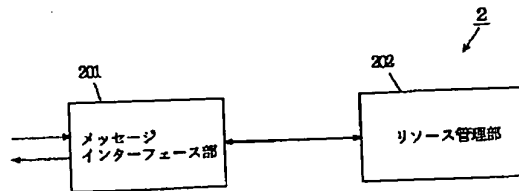


【図5】

スロット番号	予約ユーザ番号	割り当て要求番号	非優先フラグ
1	2	2	OFF
1	NO	NO	OFF
2	15	NO	OFF
...

3204, 3704...入出力部
 3205, 3705...記憶装置
 3206, 3706...MSFS
 3207, 3707...I/Oユニット
 3208, 3708...ユーザ端末
 3301, 3801...優先入出力制御部
 3302, 3802...非優先入出力制御部
 3321, 3841...コマンドキュー
 3322, 3842コマンドキュー制御部
 3323, 3843...コマンド送出部
 3324, 3844...データバッファ部
 3325...バッファ管理部
 3326, 3846...データ送出部
 3327, 3847...コマンド受付部
 3848...カウンタ
 3849...送出キュー
 4301 サーバ装置
 4302 データ記憶装置
 4310 データ送出部
 4311 データ取得部
 4312 バッファ
 4313 バッファ管理部
 4314 ポーズ制御部

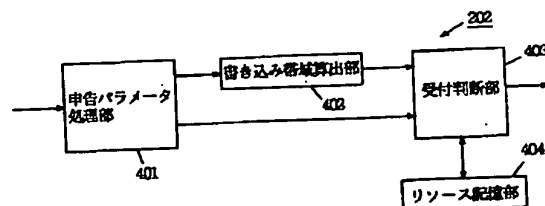
【図3】



【図6】

宛先ユニット番号	スロット番号	割り当て要求番号
1	1	12
1	2	3
2	1	NO
...

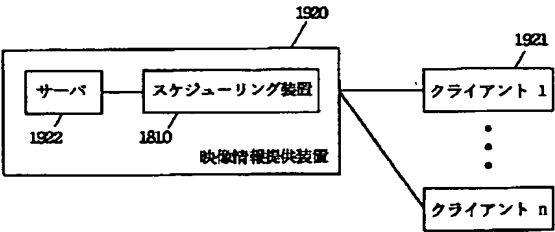
【図8】



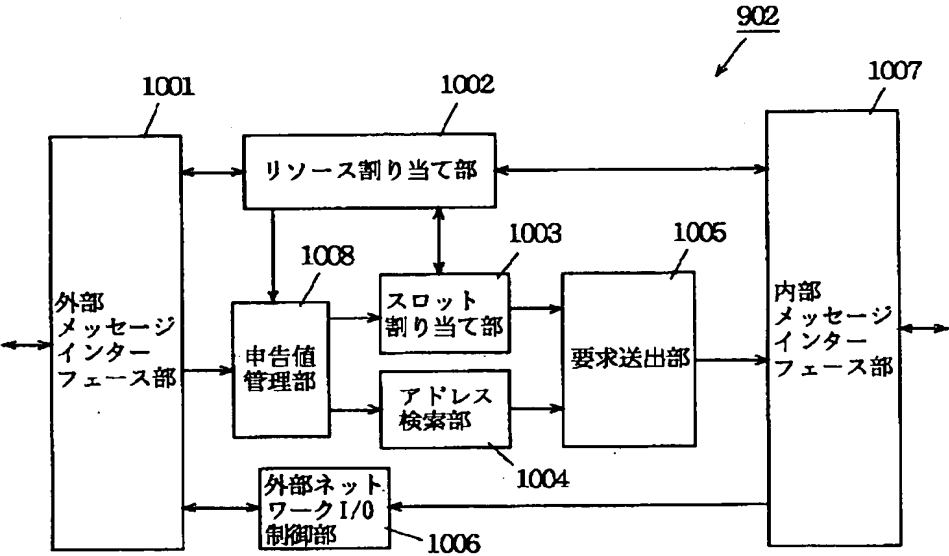
【図7】

リソース 記述子	読み出し・書き 込みの別	スロット平均 予約間隔	確保されたファイル システム内部領域	現在使用可能な ファイルシステム 内部領域
1	read	85ms	6Mbps	0
2	read	125ms	4Mbps	0
⋮	⋮	⋮	⋮	⋮

【図19】

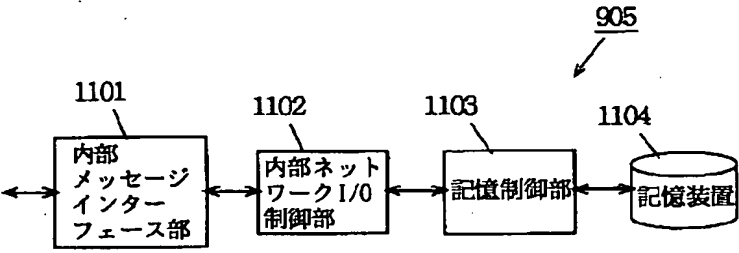


【図10】



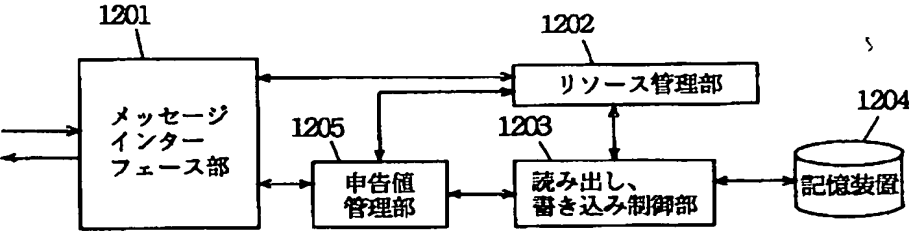
【図11】

【図21】

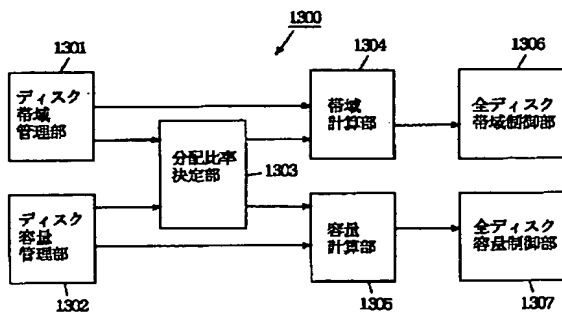


クライアント ID	T req	T ofs

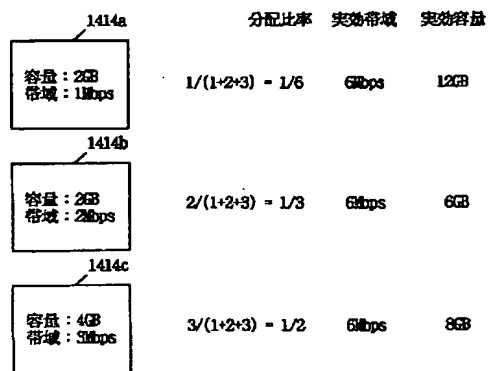
【図12】



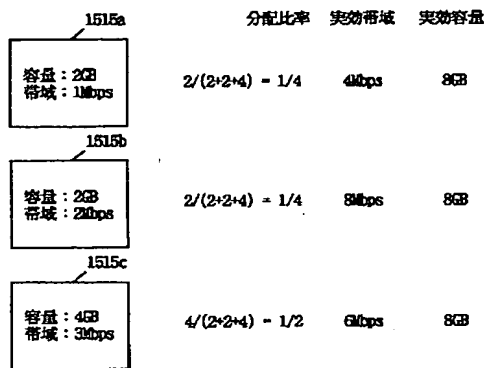
【図13】



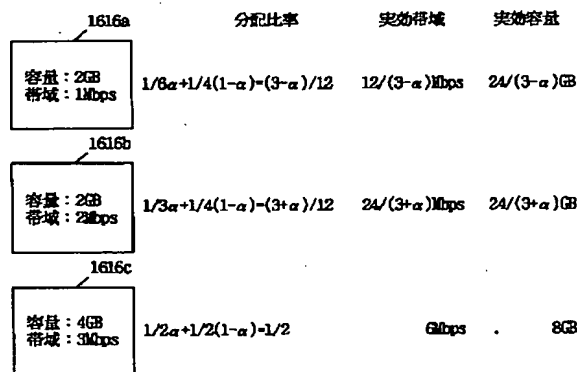
【図14】



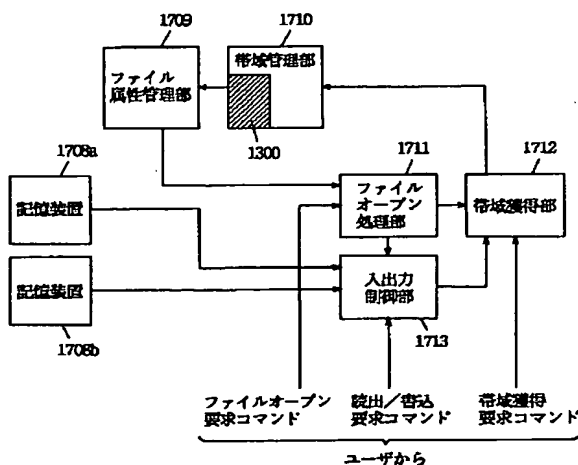
【図15】



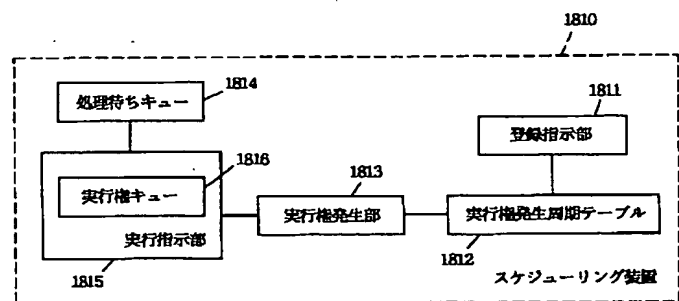
【図16】



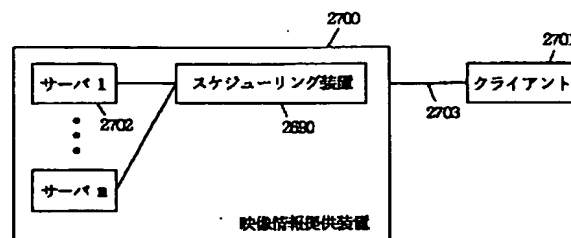
【図17】



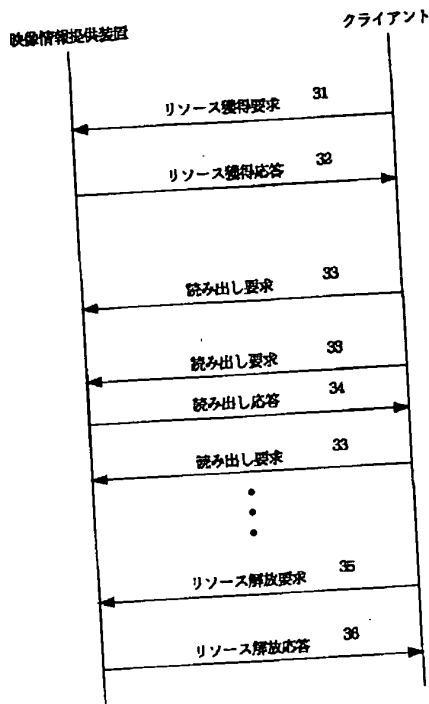
【図18】



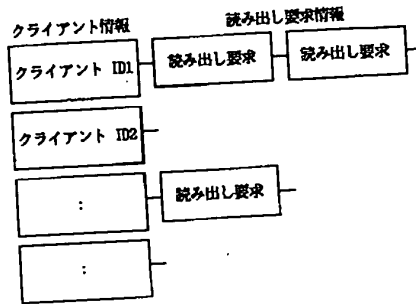
【図27】



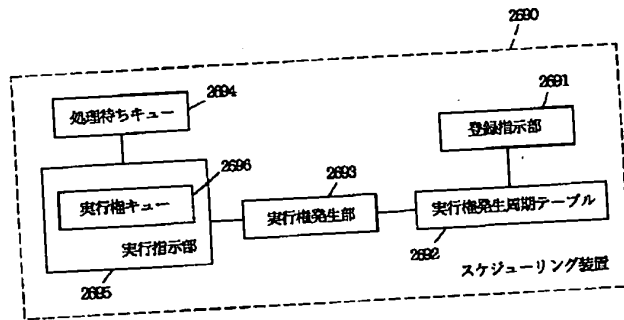
【図20】



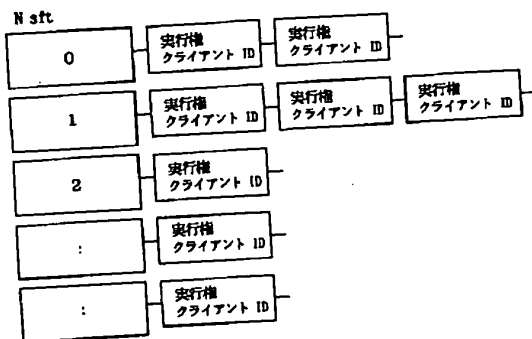
【図22】



【図26】



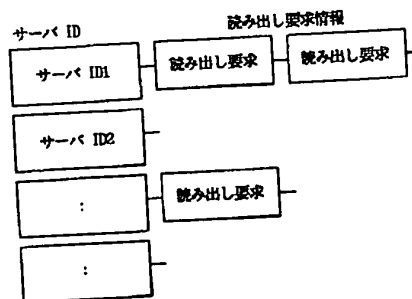
【図23】



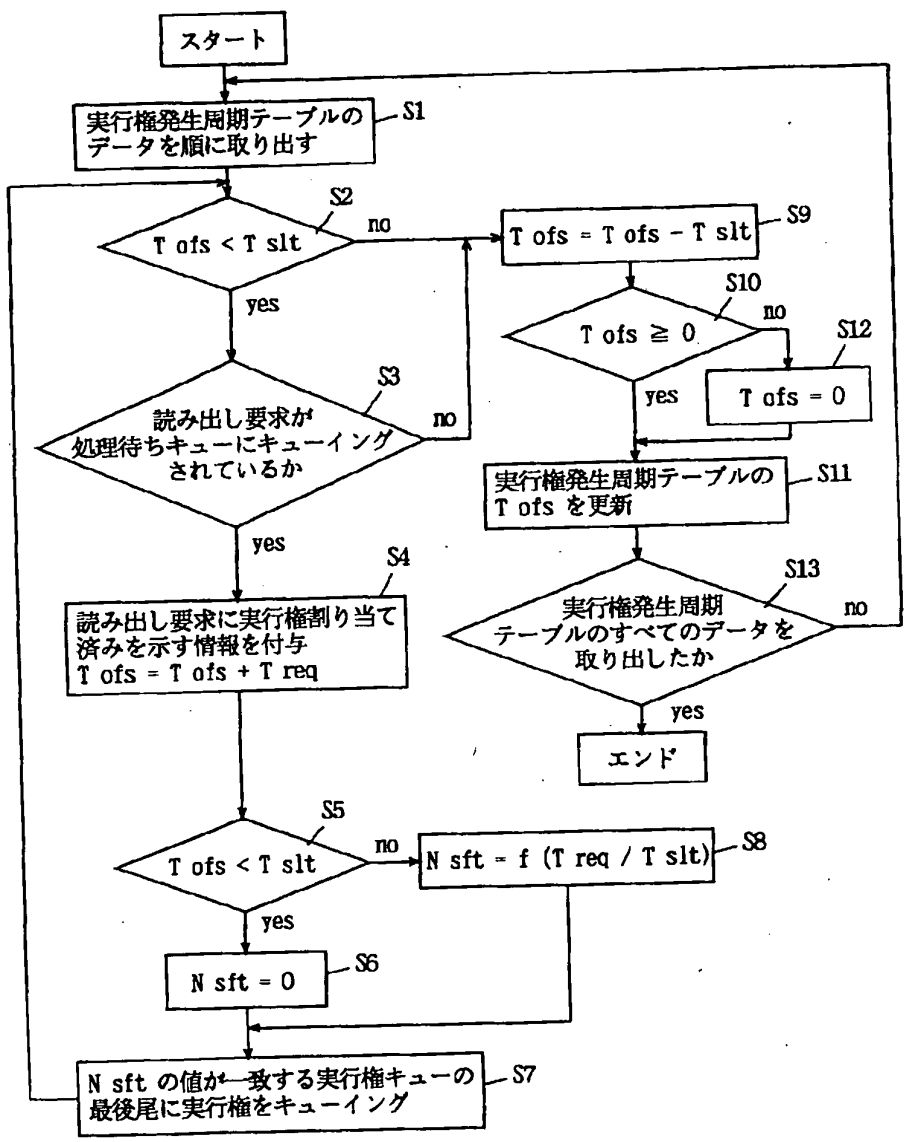
【図28】

サーバ ID	T ltr	T ofs

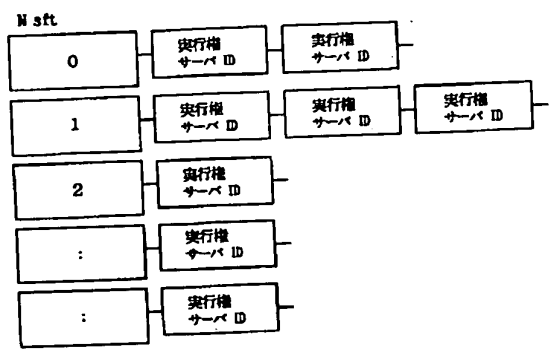
【図29】



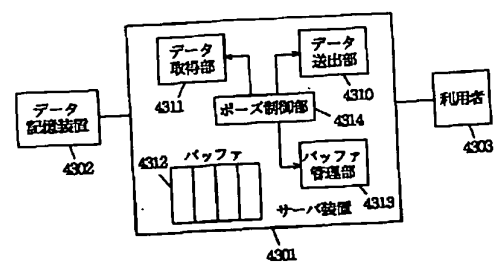
【図24】



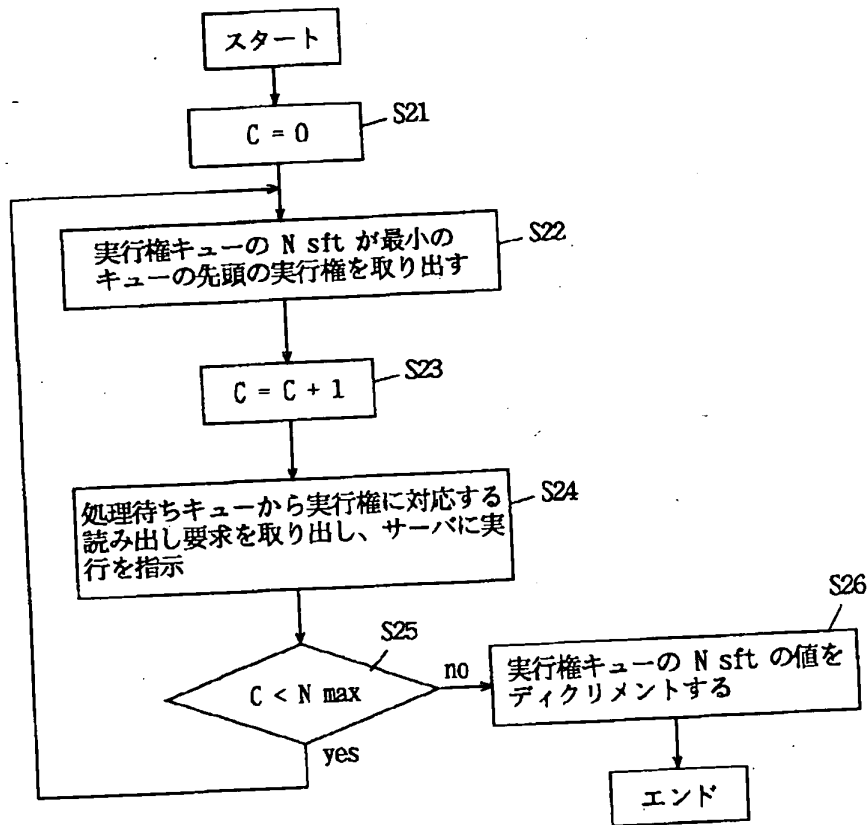
【図30】



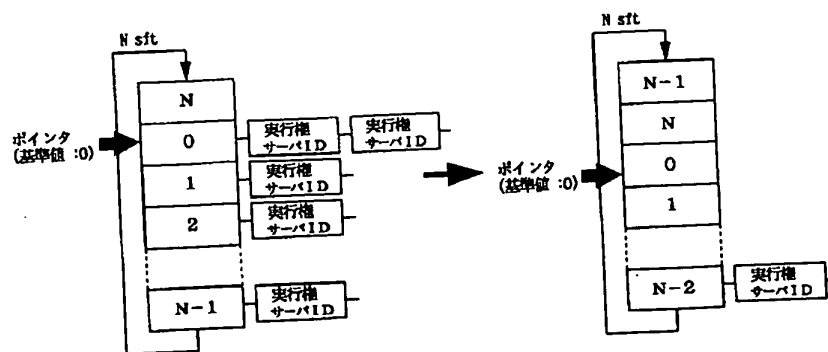
【図43】



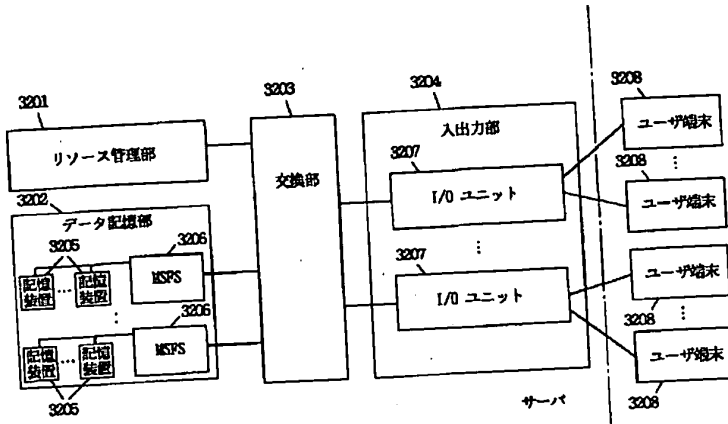
【図 25】



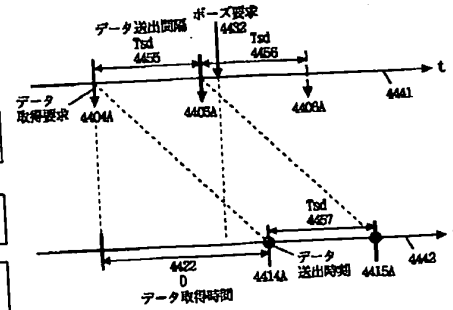
【図 31】



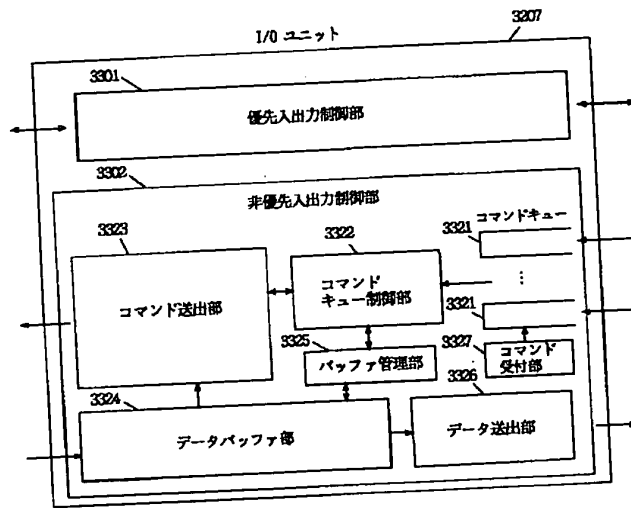
【図 32】



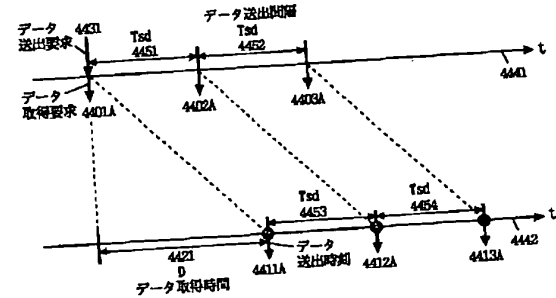
【図 45】



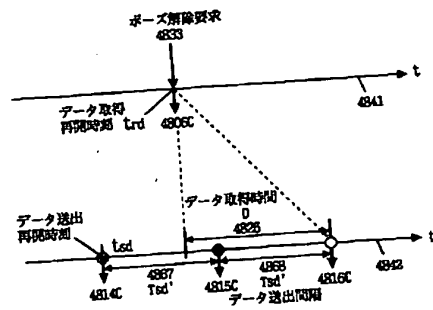
【図 33】



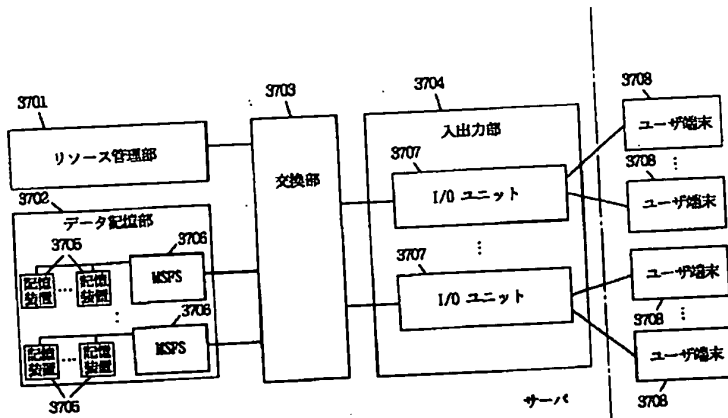
【図 44】



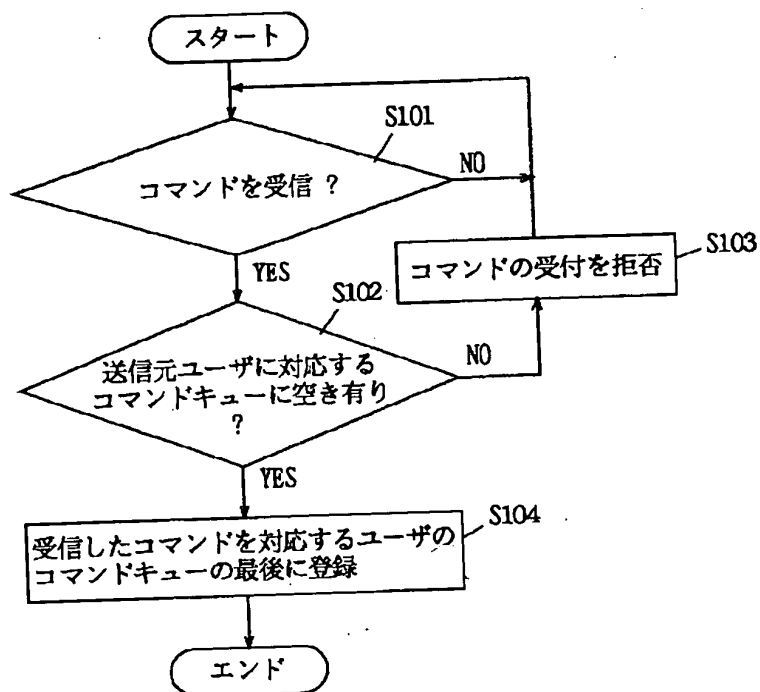
【図 48】



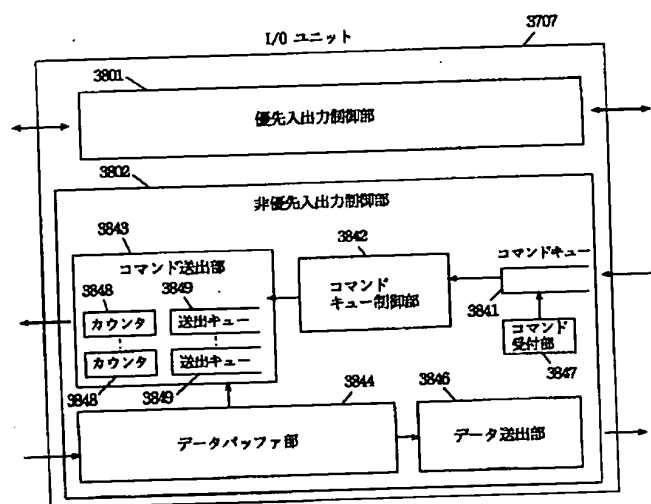
【図 37】



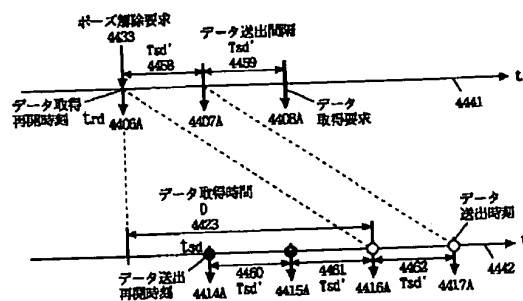
【図34】



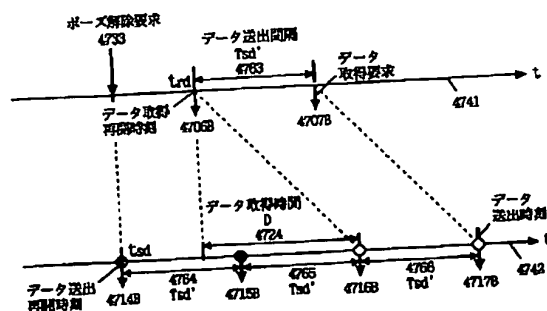
【図38】



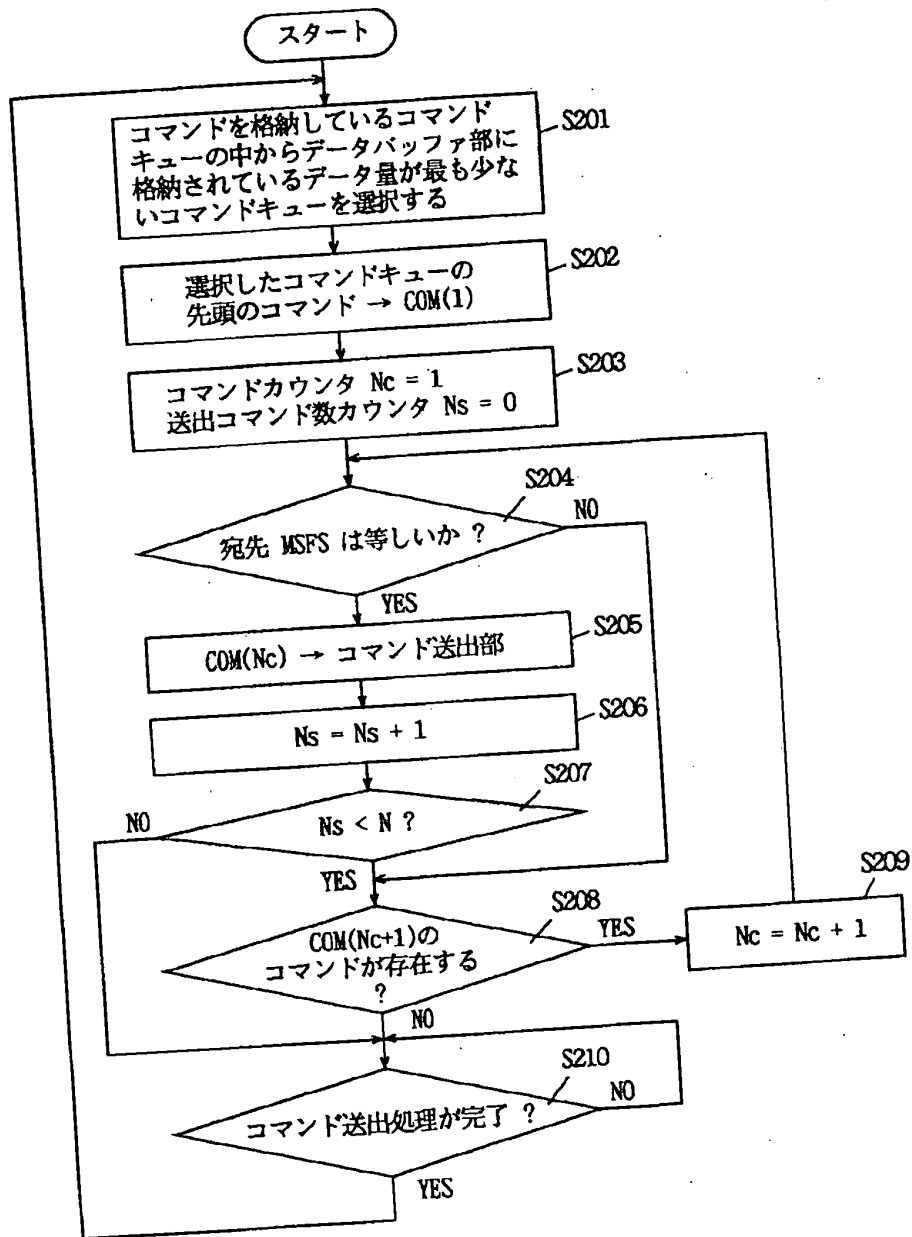
【図46】



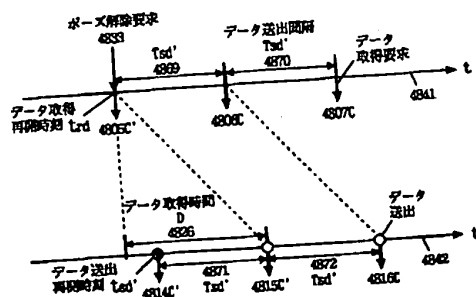
【図47】



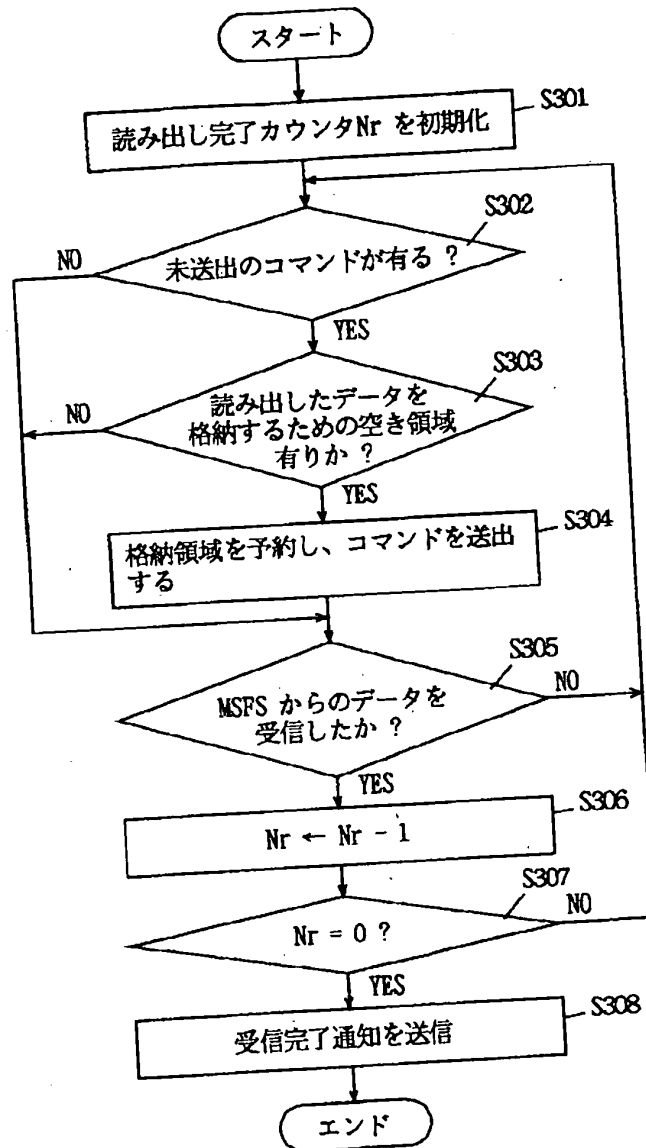
【図35】



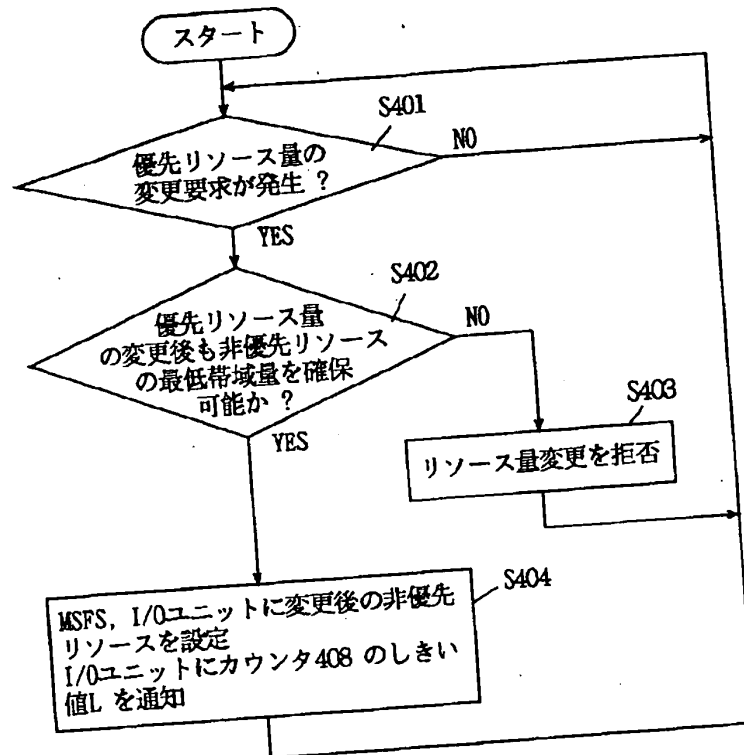
【図49】



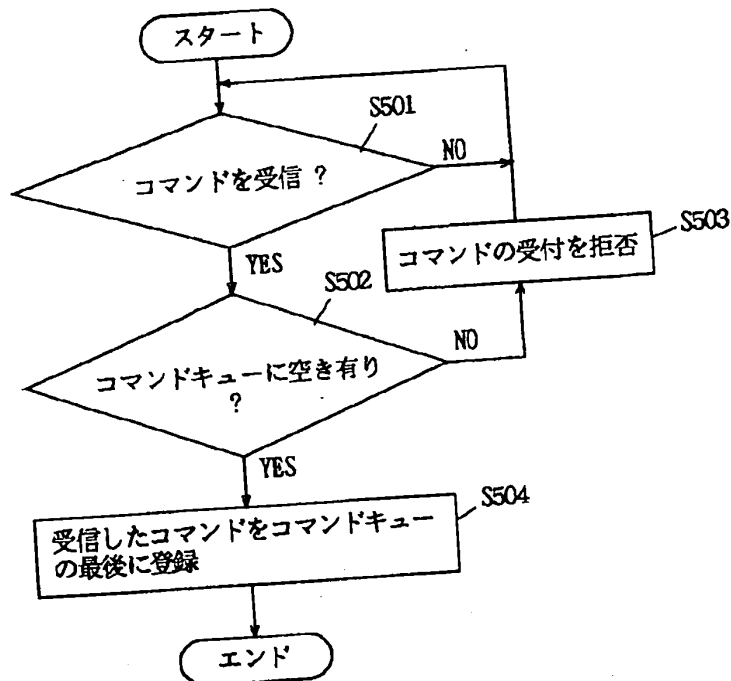
【図36】



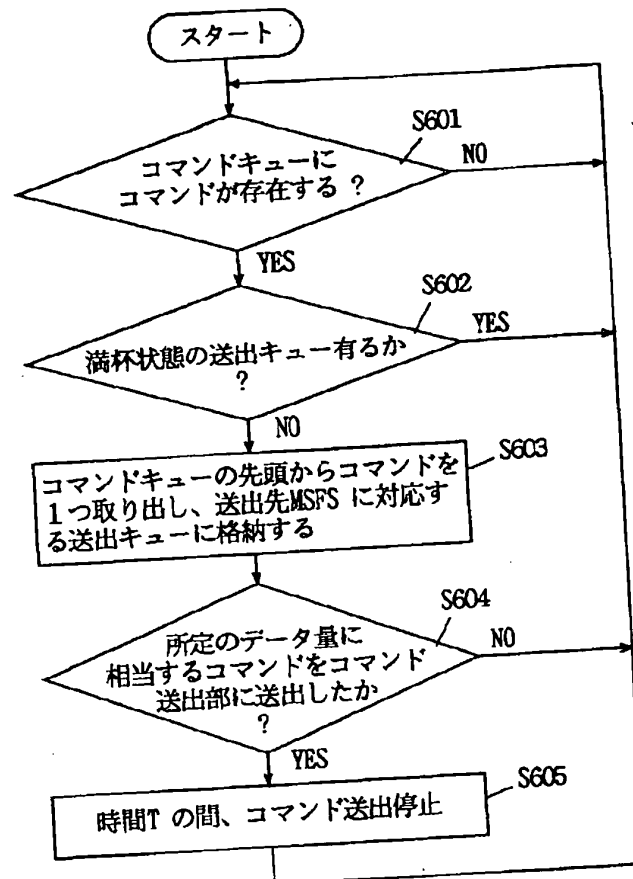
【図39】



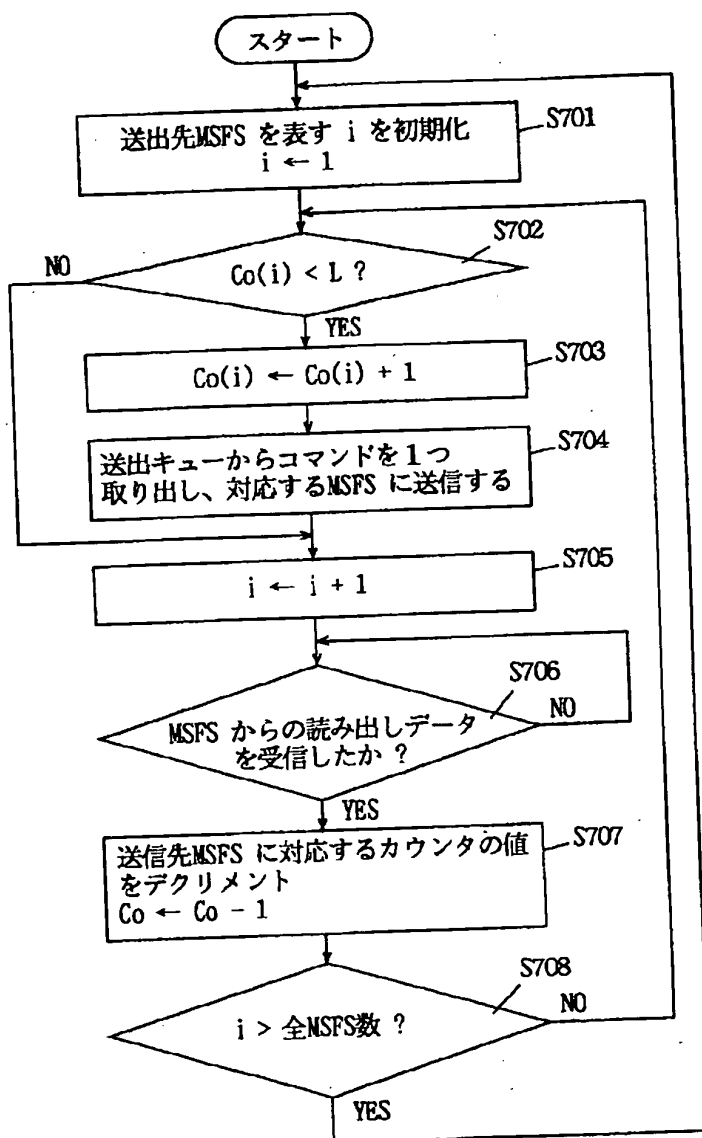
【図40】



【図41】



【図42】



フロントページの続き

(31) 優先権主張番号 特願平8-125146
 (32) 優先日 平8(1996)5月20日
 (33) 優先権主張国 日本(JP)
 (31) 優先権主張番号 特願平8-213556
 (32) 優先日 平8(1996)8月13日
 (33) 優先権主張国 日本(JP)

(72) 発明者 古門 健
 大阪府門真市大字門真1006番地 松下電器
 産業株式会社内
 (72) 発明者 荻部 朋幸
 大阪府門真市大字門真1006番地 松下電器
 産業株式会社内
 (72) 発明者 玉井 昌朗
 大阪府門真市大字門真1006番地 松下電器
 産業株式会社内